

Tipo de documento: Tesis de maestría

Master in Management + Analytics

Una aproximación al Riesgo Reputacional Bancario mediante técnicas de Text Mining

Autoría: Del Rosso, Rodrigo

Fecha de defensa de la tesis: 2023

¿Cómo citar este trabajo?

Del Rosso, R. (2023) "*Una aproximación al Riesgo Reputacional Bancario mediante técnicas de Text Mining*".
[Tesis de maestría. Universidad Torcuato Di Tella].

Repositorio Digital Universidad Torcuato Di Tella

<https://repositorio.utdt.edu/handle/20.500.13098/12028>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Creative Commons Atribución-No Comercial-Compartir Igual 2.5 Argentina (CC BY-NC-SA 2.5 AR)

Dirección: <https://repositorio.utdt.edu>



**UNIVERSIDAD
TORCUATO DI TELLA**

MASTER IN MANAGEMENT + ANALYTICS

**Una aproximación al Riesgo Reputacional
Bancario mediante técnicas de Text Mining**

TESIS

Rodrigo Del Rosso

Abril 2023

Tutor: Andrés Gago

Esta página se deja en blanco intencionalmente

Dedicatoria

A mi madre María del Carmen

In Memoriam

Esta página se deja en blanco intencionalmente

Agradecimientos

Quisiera expresar mi sincero agradecimiento a todas las personas que me han ayudado a lo largo de esta hermosa aventura de realizar esta Maestría, y que hicieron lo posible para que esta Tesis haya podido salir adelante.

En primer lugar, quiero exhibir mi agradecimiento al Director de esta tesis, el Profesor Andrés Gago por los comentarios realizados que me ayudaron a enriquecer el trabajo presentado.

En especial, al amor de mi vida Lara Woelflin, por su incondicional apoyo en cada proyecto que emprendo y por impulsarme a mejorar y progresar. Y a Pepa por ser una gran compañera en cada rato libre dedicado a este trabajo.

A mis compañeros de cursada, en particular a mi grupo de estudio constituido desde el inicio: Gustavo Levinis, Guillermo Klein, Esteban Bertuccio, Andrés Stocks y Fernando Spannaus con quienes compartí muchísimas horas de estudio, viajes a la Universidad en distintos medios de transportes, la realización de muchísimos trabajos prácticos, intercambios de ideas y una linda amistad que mantengo con ellos. Además, agradecer a otros compañeros de la misma camada con quienes compartí muchas clases e intercambio de opiniones en los *breaks*: Sheila Garade, Nicolás García Aramouni, y Aldo Escobar.

Asimismo, agradecer a otros compañeros de las siguientes cohortes que fueron de apoyo emocional en el trayecto final de la cursada: María Paula Laspeñas y Oscar David Vargas Duarte.

Al personal administrativo de la Maestría, a Delfina Cidre y Maia Vachiega que siempre estuvieron a plena disposición ante cualquier pedido o requerimiento realizado y a la dirección a cargo del Profesor Juan José Miranda Bront.

A cada uno de los excelentes profesores de esta casa de estudios que tuve la oportunidad de conocer, que me despertaron interés y pasión por los contenidos impartidos en el posgrado.

A mis amistades, Martin Masci, Nuria Hernández, Agustina Galgano, Guido Bavosa Castela, Federico Tassara, Juan Ignacio Serrats, Hernán Escobar, Matias Saffigna, Natalia Barral, Mauro Speranza, Juan Cruz Gravina y Lorena Toledo por el sostén en distintos momentos de mi vida. En especial a Martin, que tuvo la gentileza de releer el trabajo en profundidad y darme el impulso final para terminarla. Sin su contribución este trabajo no hubiese sido posible presentarlo.

A quienes se tomaron el tiempo de releer cada capítulo para brindarme sus correcciones y comentarios, en especial a mi director Andrés, a mi pareja Lara y a mis amigos, Agustina, Gustavo, Martin y Paula.

Por último, agradecer a mi familia y en especial a la memoria de mi madre, que está siempre presente en cada momento de mi vida. Esta tesis está dedicada especialmente a su memoria.

Esta página se deja en blanco intencionalmente

Resumen

Una aproximación al Riesgo Reputacional Bancario mediante técnicas de Text Mining

Rodrigo Del Rosso

rodrigodelrosso@gmail.com

El modelo desarrollado se utilizará para gestionar el riesgo reputacional de la empresa, reflejando el importe de Capital que tendrá que mantener reservado para hacer frente a los eventuales efectos adversos originados por una percepción negativa por parte de todos los interesados en la organización. Se espera que este modelo permita periódicamente determinar el nivel de asignación óptima de capital para mitigar efectos adversos que pudiesen dañar la solvencia de la compañía, a partir de la información suministrada por el mercado, el contexto macroeconómico local y la percepción de los distintos participantes.

Ante la falta de regulación específica, esta tesis es una propuesta para mitigar el riesgo reputacional de las Entidades Financieras y por lo tanto, da respuesta a una pretensión de la autoridad monetaria de gestionar el riesgo reputacional. En este trabajo se desarrolla un modelo econométrico que incorpora datos provenientes de las percepciones que expresan los usuarios a través de opiniones públicas a partir del empleo de técnicas de *Text Mining*. Para eso se utilizaron datos públicos extraídos de las opiniones de usuarios de la red social *Twitter* y las cotizaciones de los precios de aquellos Bancos que cotizan en el BYMA (Bolsa y Mercados Argentinos). Además, se empleó información financiera propia del rol de la empresa cotizante dentro del índice bursátil S&P Merval.

Dentro del proceso interno de evaluación de la adecuación del capital (ICAAP) que deben realizar las Entidades Financieras en el marco de su política de Gestión de Riesgos, el Banco Central de la República Argentina (BCRA), establece que dichas instituciones deben desarrollar metodologías para medir en la forma más precisa posible los efectos del riesgo reputacional en términos de otros tipos de riesgo a los cuales podría estar la entidad expuesta, incluyendo para ello escenarios de riesgo reputacional en sus pruebas de tensión regulares a los fines de evitar efectos reputacionales adversos y mantener la confianza del mercado. En la literatura se mencionan algunos enfoques econométricos con variables financieras para dimensionar el Capital Económico que deben destinar para hacer frente al mencionado riesgo. Asimismo, se hace alusión a que este riesgo puede asimilarse al Operacional, y no existe un consenso sobre la forma de cuantificarlo.

Palabras Claves: Mercados financieros y Macroeconomía, Instituciones Financieras y Servicios, Riesgo Financiero, Riesgo Reputacional, Evaluación de Riesgos.

Clasificación JEL: E44, G2, G210

Abstract

An approach to Reputational Banking Risk using Text Mining techniques

Rodrigo Del Rosso

rodrigodelrosso@gmail.com

The model developed will be used to manage the company's reputational risk, reflecting the amount of Capital that it will have to keep reserved to face possible adverse effects caused by a negative perception on the part of all those interested in the organization. This model is expected to periodically determine the optimal capital allocation level to mitigate adverse effects that could damage the solvency of the company, based on the information provided by the market, the local macroeconomic context and the perception of the different participants.

Given the lack of specific regulation, this thesis is a proposal to mitigate the reputational risk of Financial Institutions and therefore, responds to a claim by the monetary authority to manage reputational risk. In this work, an econometric model is developed that incorporates data from the perceptions expressed by users through public opinions from the use of Text Mining techniques. For this, public data extracted from the opinions of users of the Twitter social network and the price quotes of those Banks listed on the BYMA (Bolsa y Mercados Argentinos) were used. In addition, financial information from the role of the listed company within the S&P Merval stock index was used.

Within the internal capital adequacy assessment process (ICAAP) that Financial Institutions must carry out within the framework of their Risk Management policy, the Central Bank of the Argentine Republic (BCRA) establishes that said institutions must develop methodologies to measure as accurately as possible the effects of reputational risk in terms of other types of risk to which the entity could be exposed, including reputational risk scenarios in its regular stress tests in order to avoid adverse reputational effects and maintain market confidence. Some econometric approaches with financial variables are mentioned in the literature to measure the Economic Capital that must be allocated to face the aforementioned risk. Likewise, reference is made to the fact that this risk can be assimilated to Operational risk, and there is no consensus on how to quantify it.

Keywords: Financial Markets and Macroeconomics, Financial Institutions and Services, Financial Risk, Reputational Risk, Risk Assessment.

JEL Classification: E44, G2, G210

Índice

Dedicatoria	3
Agradecimientos	5
Resumen	7
Abstract	8
Introducción	13
Pregunta de investigación y objetivos	14
Capítulo 1: Aspectos normativos y técnicos de la medición del Riesgo Reputacional	16
1.1. Nociones sobre los Riesgos en Entidades Financieras	16
1.2. Sobre el Riesgo Reputacional	18
1.3. Acerca de la Medición del Riesgo Reputacional	23
1.4. En búsqueda de la Solvencia	28
Resumen de Capítulo	31
Capítulo 2: Text Mining aplicado a la medición del Riesgo Reputacional	33
2.1. Sobre el análisis de sentimiento	33
2.2. Acerca de la medición del Sentimiento y el modelo de Polarización	35
2.3. Los modelos de NLP en el análisis de sentimiento aplicado en contextos económicos y financieros	42
Resumen de Capítulo	44
Capítulo 3: Aplicación de Text Mining para medir el Riesgo Reputacional	45
3.1. Análisis Exploratorio de Datos	45
3.2. Aplicación de Sentiment Analysis al cálculo del Capital Económico	51
3.2.1. Metodología propuesta	51
3.2.2. Text Mining sobre los datos	52
3.2.3. Resultados del modelo estimado	57
3.2.4. Sobre el Backtesting del modelo	60
3.3. Recomendaciones de Gestión	64
Resumen de Capítulo	65
Conclusiones y Futuras Investigaciones	67

Referencias Bibliográficas	69
Bibliografía	72
Apéndices	73
Datos	73
Códigos	73
Tablas de Formas Funcionales	74
Gráficos	76
Análisis de Normalidad sobre los retornos diarios	76
Correlación de Retornos Diarios	83
Análisis de Outliers de Retornos Diarios	85
Evolución de Sentimiento Diario	92

Índice de Figuras

1	Mapa de Riesgos de un Banco	17
2	KPMG - Informe 2019	21
3	KPMG - Informe 2019	22
4	Distribución de Probabilidad de Pérdidas Potenciales	30
5	Diagrama de Flujo del Modelo	35
6	Modelo de <i>Bags-of-Words</i>	36
7	Evolución de la cotización de cada Acción	47
8	Evolución de la cotización de cada Acción	48
9	Evolución de los Retornos Acumulados	49
10	Histogramas de los Retornos Diarios	50
11	Cantidad de Menciones por hora	53
12	Densidad por fecha y por hora	54
13	Sentimiento Diario de Menciones	55
14	Histograma del Sentimiento Diario	56
15	Modelo de Regresión Lineal Simple	58
16	Información de mercado de la Acción BRIO.BA	60
17	Sentimiento Diario de Menciones	61
18	Cantidad de Menciones Negativas por Hora	62
19	Histograma del Sentimiento Diario	63
20	Formas Funcionales del Peso Local	74
21	Formas Funcionales del Peso Global	75
22	Formas Funcionales del Factor de Normalización	75
23	Diagnóstico de Normalidad - Retornos Hipotecario	76
24	Diagnóstico de Normalidad - Retornos Valores	77
25	Diagnóstico de Normalidad - Retornos Macro	78
26	Diagnóstico de Normalidad - Retornos Patagonia	79
27	Diagnóstico de Normalidad - Retornos BBVA Francés	80
28	Diagnóstico de Normalidad - Retornos Santander Argentina	81
29	Diagnóstico de Normalidad - Retornos Galicia	82
30	Correlación - Retornos Diarios	83
31	Correlación - Retornos Diarios	84
32	Outliers - Banco Hipotecario	85
33	Outliers - Banco de Valores	86

34	Outliers - Banco Macro	87
35	Outliers - Banco Patagonia	88
36	Outliers - Banco BBVA Francés	89
37	Outliers - Banco Santander Argentina	90
38	Outliers - Banco Galicia	91
39	Sentimiento Diario de Menciones Negativas	92
40	Sentimiento Diario de Menciones Negativas	93
41	Sentimiento Diario de Menciones Negativas	94

Índice de Tablas

1	Denominación en BYMA	46
2	Ranking de Activos - Sistema Financiero	46
3	Resumen de medidas descriptivas	50
4	Resumen de medidas descriptivas	57
5	Estimación de Especificaciones Propuestas	59
6	Medidas Descriptivas	64

Esta página se deja en blanco intencionalmente

"...El tiempo a mí me puso en otro lado, tendré que hacer lo que es no debido, tendré que hacer el bien y hacer el daño. No olvides que el perdón es lo divino, y errar, a veces, suele ser humano..."

Fito Páez, *Al lado del camino*

Introducción

El presente trabajo constituye una aproximación a la medición del Riesgo Reputacional de Entidades Financieras. Este riesgo es uno de los más difíciles de medir dentro del proceso interno de evaluación de la adecuación del capital (ICAAP), porque no existe un consenso en la industria sobre la forma de cuantificarlo. El proceso de autoevaluación del capital se debe realizar en el marco de la política integral de Gestión de Riesgos según la normativa definida por el Banco Central de la República Argentina (BCRA), donde el área de Gestión de Riesgos debe evaluar la exposición a cada uno de los riesgos a los que se exponen¹.

En palabras de Freixas *et al.* (1997, p. 3) “un banco es una institución cuyas operaciones habituales consisten en conceder préstamos y recibir depósitos del público.”. Por lo cual, factores como la confianza y la credibilidad y, consecuentemente, el Riesgo Reputacional, resultan de vital importancia en este tipo de organizaciones. Por lo tanto, la pérdida de reputación organizacional puede generar daños tan profundos en las entidades como para ocasionarles la quiebra.

Por tal motivo, resulta conveniente destacar que en los últimos años se ha detectado un sustancial incremento de este riesgo por numerosas causas como enumera Villanueva (2020),

"Los riesgos reputacionales han aumentado en los últimos años para las empresas e instituciones financieras por efecto, principalmente, de la aparición de las redes sociales; la creciente vigilancia y sensibilidad respecto a la corrupción y las conductas fraudulentas e ilícitas; y la demanda de unos servicios cada vez de más calidad e igualitarios por parte de los nuevos movimientos sociales y ciudadanos. (...)" (Villanueva, 2020, p.1)

Siguiendo los lineamientos definidos por el BCRA, las Entidades Financieras deben desarrollar metodologías para medir los efectos del riesgo reputacional en términos de otros tipos de riesgo a los cuales podrían estar expuestas. Entre algunas de las técnicas empleadas se incluye la utilización de escenarios de riesgo reputacional en sus pruebas de estrés a los fines de evitar efectos adversos y mantener la confianza del mercado.

Asimismo, no existe un consenso sobre la forma de cuantificarlo (Soprano *et al.*, 2009) y se lo suele asimilar al riesgo operacional. Por ejemplo, algunos autores emplean un enfoque econométrico basado en la utilización de un índice reputacional, que se determina en

¹La normativa es la Comunicación “A” 5398 - Lineamientos para la Gestión de Riesgos en las Entidades Financieras. El BCRA es el órgano de supervisión y control de la República Argentina.

base a un análisis de sentimiento sobre relevamientos de encuestas (Mitic, 2018). Estos resultados se utilizan para aproximar la distribución de pérdidas mediante una simulación de eventos reputacionales y de esta manera, se estima el capital que debería reservar la organización para hacer frente a eventuales efectos adversos originados por una percepción negativa².

Ante la falta de regulación específica sobre el Riesgo Reputacional en las Entidades Financieras, este trabajo es una propuesta para mitigarlo y por lo tanto da respuesta a la pretensión de la autoridad monetaria sobre el cálculo del capital que deberían reservar para cubrir los eventos reputacionales (Capital Económico). Este modelo debe ser utilizado por aquellas áreas de Gestión Integral de Riesgos de Bancos. A tales fines, en este trabajo se realiza una estimación de un modelo econométrico incluyendo como regresor a un indicador de la reputación mediante el empleo de técnicas de *Text Mining*.

La tesis se estructura en tres capítulos. En el primero de ellos se realizará un análisis del marco normativo regulatorio para la medición del Riesgo Reputacional en Entidades Financieras. En primer lugar, se estudiarán las nociones sobre los riesgos en las Entidades Financieras documentadas en la literatura y en las normativas. En segundo lugar, se analizarán los componentes que lo constituyen, diversas definiciones y aquellos factores que lo determinan. Finalmente, se esbozarán aspectos tradicionales de medición utilizados para su cuantificación.

Para abordar la medición de este riesgo en el mercado financiero, en el segundo capítulo se ahondará en la utilización de técnicas de *Text Mining*, en particular de Análisis de Sentimiento para medir el Riesgo Reputacional de Entidades Financieras.

Por último, en el tercer capítulo, se realizará la estimación del Capital que debería reservar el Banco Santander Argentina S.A. para hacer frente ante eventuales efectos adversos originados por una percepción negativa de la organización.

Pregunta de investigación y objetivos

A partir de lo expuesto previamente, surge un problema de investigación formulado a partir de la siguiente pregunta, ¿Cómo evaluar si las técnicas de *Text Mining* pueden ser útiles para la medición del riesgo reputacional y cómo aplicarlas para determinar un capital de reserva? y para responderla se plantea el siguiente objetivo general,

- Estimar el cálculo del Capital Económico por Riesgo Reputacional mediante la utilización de técnicas de *Text Mining*.

A partir de este objetivo general, se pueden plantear tres objetivos específicos que serán abordados cada uno en un capítulo distinto de la tesis:

1. Analizar los aspectos normativos y técnicos para la medición del Riesgo Reputacional.
2. Profundizar en las aplicaciones de *Sentiment Analysis* con datos públicos extraídos de *Twitter*.

²Al enfoque de simulación para estimarlo se lo denomina LDA - *Loss Distribution Approach*

3. Aplicar la técnica de *Sentiment Analysis* en la estimación del Capital Económico por este riesgo.

"...Dar es dar, es solamente una manera de andar..."

Fito Páez, *Dar es dar*

Capítulo 1: Aspectos normativos y técnicos de la medición del Riesgo Reputacional

En este capítulo se analiza el marco normativo regulatorio para la medición del Riesgo Reputacional en Entidades Financieras de Argentina y se brinda un análisis de los principales aspectos normativos y técnicos que un profesional que se desempeña en un área de Gestión de Riesgos debe manejar para gestionarlo en forma más adecuada.

Para cumplimentar con dicho objetivo, en primer lugar, se pasa revista a las nociones sobre los riesgos en las Entidades Financieras documentadas en la literatura, en las normativas locales y en las recomendaciones realizadas por el Comité de Supervisión Bancaria de Basilea. En segundo lugar, se focaliza en el Riesgo Reputacional, los componentes que lo constituyen, diversas definiciones y los determinantes del mismo. Finalmente, se esbozan aspectos tradicionales de medición utilizados para cuantificarlo.

1.1. Nociones sobre los Riesgos en Entidades Financieras

Existe Riesgo cuando puede ocurrir un evento que tiene consecuencias financieras adversas. La administración o gerenciamiento de los riesgos financieros (***Financial Risk Management***) es el proceso de identificación, medición y control de los mismos. Jorion (2011) estipula que existen algunos riesgos que pueden ser medidos razonablemente y otros que, siendo importantes, no son susceptibles de una medición formal,

“Some risks can be measured reasonably well. For those, risk can be quantified using statistical tools to generate a probability distribution of profits and losses. Other risks are not amenable to formal measurement but are nonetheless important. The function of the risk manager is to evaluate financial risks using both quantitative tools and judgment (...)” (Jorion, 2011, p.74)

Según el BCRA, las Entidades Financieras deben contar con un proceso integral para la Gestión de Riesgos para identificar, evaluar, seguir, controlar y mitigar *todos* los riesgos significativos. Además debe ser proporcional a la dimensión e importancia económica del Banco y tener en cuenta la naturaleza y complejidad de las operaciones realizadas (BCRA, 2013).

Algunos autores hacen referencia a tres fuentes principales de riesgos que afectan a los Bancos: de Crédito, de Tasa de Interés y de Liquidez (Freixas *et al.*, 1997). Sin embargo, existen otros factores que pueden afectarlos (Bessis, 2015), como por ejemplo el Riesgo de Mercado, el Operacional, el Reputacional, el Legal, el Estratégico, y el de Fraude mediante operaciones de Lavado de Activos y Financiamiento del Terrorismo (LA/FT), entre otros

riesgos. En el siguiente gráfico se exhibe un mapa de relación de los distintos riesgos que un gerente gestiona diariamente. En este trabajo se propone una metodología para medir el riesgo reputacional.

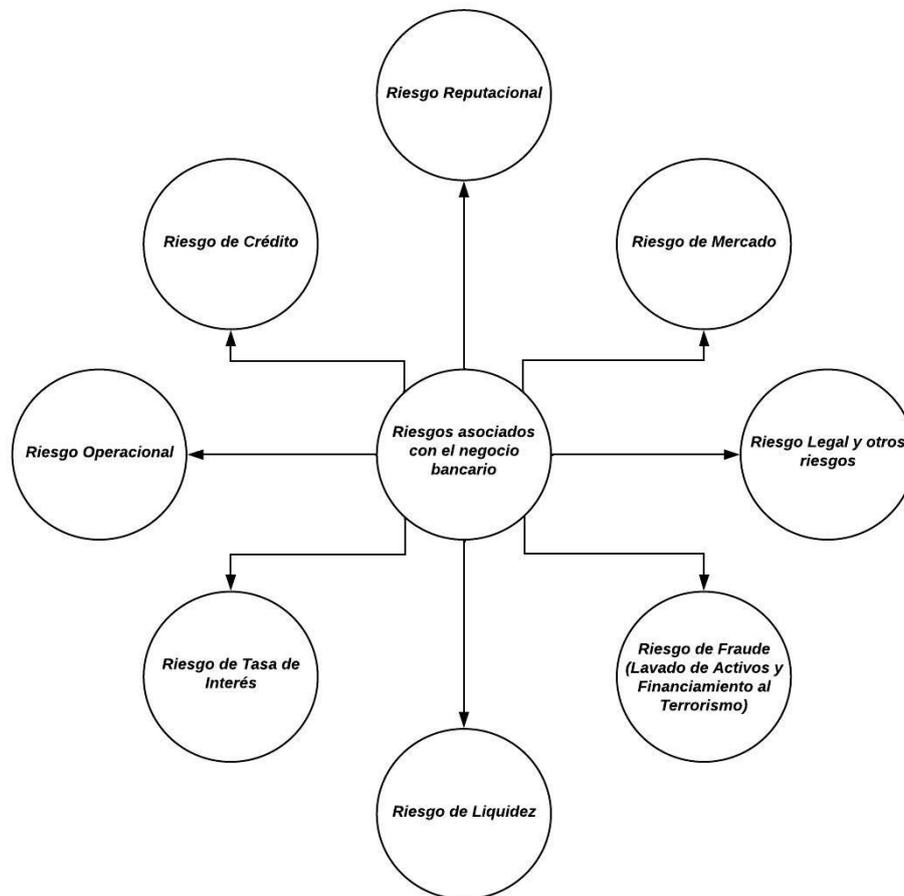


Figura 1. Mapa de Riesgos de un Banco

El Riesgo de Crédito hace referencia a la posibilidad de sufrir pérdidas por el incumplimiento de las obligaciones pactadas en los contratos que un deudor o contraparte efectúe y se puede encontrar presente tanto en las operaciones dentro como fuera del Balance. Asimismo, incluye al Riesgo País y al Riesgo de Crédito de Contraparte (BCRA, 2013). Novello (2008) lo define como *“la probabilidad de experimentar una pérdida, ya sea porque un deudor incurre en default o porque se deteriora su calidad crediticia.”*

El Riesgo de Tasa de Interés forma parte del Pilar II (Proceso de Examen Supervisor) del Marco de Basilea II³. También se encuentra contenido en las directrices establecidas por el Comité de Supervisión Bancaria de Basilea en el documento consultivo denominado *“Principles for the Management and Supervision of Interest Rate Risk”* de 2004. En la normativa regulatoria de Argentina se menciona en la Comunicación “A” 5.398 del BCRA sobre *“Lineamientos para la Gestión de Riesgos en las Entidades Financieras”*.

El Riesgo de Liquidez se vincula con la capacidad que tienen las entidades de fondear el incremento de los Activos y de cumplir con sus obligaciones a medida que éstas se hacen exigibles, sin incurrir en pérdidas significativas (BCRA, 2013). Asimismo, se puede

³El Riesgo de Tasa de Interés también conocido como Riesgo Estructural de Balance.

desagregar en dos vertientes, en fondeo o mercado. La primera, se presenta cuando no se puede cumplir de manera eficiente con los flujos esperados e inesperados, actuales y futuros, y con las garantías sin afectar las operaciones diarias o la condición financiera. Mientras que la segunda, se da cuando una Entidad no puede deshacer o compensar una posición a precio de mercado, debido a que los activos que la componen no cuentan con un mercado secundario suficiente o por alteraciones del mismo (BCRA, 2013).

El Riesgo de Mercado se refiere a la posibilidad de que la Entidad sufra pérdidas tanto en posiciones dentro como fuera del Balance producto de fluctuaciones contrarias en los precios de mercado de diversos activos. Está compuesto por los riesgos inherentes a las Acciones, a los instrumentos financieros cuyo valor depende de las Tasas de Interés y de otros instrumentos considerados en la Cartera de Negociación (*Trading Book*) y el riesgo de Moneda dado tanto por operaciones dentro como fuera del Balance.

El Riesgo Operacional - concepto que incluye el riesgo legal y excluye el riesgo estratégico y de reputación-, se refiere a las pérdidas resultantes por la falta de adecuación o fallas en los procesos internos, en la actuación del personal o en los sistemas, así como de aquellas que sean producto de eventos externos. El Riesgo Legal comprende, entre otros aspectos, la exposición a sanciones, penalidades u otras consecuencias económicas y de otra índole por incumplimiento de normas y de obligaciones contractuales (BCRA, 2013).

El Riesgo Estratégico es aquel procedente de una estrategia de negocios inadecuada o de un cambio adverso en las previsiones, parámetros, objetivos y otras funciones que respaldan esa estrategia. A pesar de lo complejo de su estimación, a los efectos de realizar una adecuada gestión de este riesgo, la entidad financiera debe i) considerar el enfoque estratégico, el plan de negocios y los requerimientos futuros de capital cuando evalúe el objetivo de suficiencia del capital en función del riesgo e ii) incluir con claridad en el plan estratégico cuál es el nivel deseado, las necesidades de capital y cuáles son las fuentes externas de capital disponibles.

Por último, el Riesgo de Fraude por Lavado de Activos y Financiamiento del Terrorismo se define como la posibilidad de que las Entidades Financieras sufran pérdidas producto de ser utilizadas por terceros con objetivos criminales vinculados al Lavado de Activos y al Financiamiento del Terrorismo (Res. 30/2017 – UIF).

1.2. Sobre el Riesgo Reputacional

La Reputación es un activo clave para cualquier organización que se sustente en la confianza como ocurre principalmente con las Entidades Financieras. A pesar de su importancia, la cantidad de estudios relacionados en la industria financiera sigue siendo limitada. En la actualidad, no existe una definición oficial que sea aceptada en forma universal por todos los supervisores (Soprano *et al.*, 2009).

Según BCRA (2013),

“se define al riesgo reputacional como aquel que está asociado a una percepción negativa sobre la entidad financiera por parte de los clientes, contrapartes, accionistas, inversores, tenedores de deuda, analistas de mercado y otros participantes del mercado relevantes que afecta adversamente la capacidad de la entidad financiera para mantener relaciones comerciales existentes o

establecer nuevas y continuar accediendo a fuentes de fondeo -tales como en el mercado interbancario o de titulización-(...)”. (BCRA, 2013, p.135)

Esta definición es la misma que la enunciada por el Comité de Supervisión Bancaria de Basilea,

“risk arising from negative perception on the part of customers, counterparties, shareholders, investors, debt-holders, market analysts, other relevant parties or regulators that can adversely affect a bank’s ability to maintain existing, or establish new, business relationships and continued access to sources of funding” .(BIS, 2009,p.19)

Es importante destacar que la reputación adquiere una especial importancia en la industria financiera por la existencia de información asimétrica, por la transformación cualitativa realizada en los Activos (*Qualitative-Asset-Transformation*) por los Bancos y por el suministro de servicios de pago y de riesgos que crean un Riesgo Sistémico (Fiordelisi *et al.*, 2011). Además, puede afectar los pasivos de la entidad, debido a que la confianza del público y la capacidad de la entidad de captar fondos están fuertemente vinculadas con su reputación. Esto puede conducir a otorgar un respaldo implícito, incurriendo en riesgos de crédito, liquidez, mercado y legal, con posibles incidencias negativas en sus resultados, en la liquidez y en el capital regulatorio.

Este riesgo presenta ciertas similitudes con el Operacional, y a veces se lo suele modelar en forma similar. Por lo tanto, eventos operativos tales como fraudes, interrupción de los sistemas, entre otros tipos de sucesos podrían tener un impacto en la reputación de la organización. Por tal motivo, las instituciones financieras deben controlar y gestionar en forma integral ambos riesgos (Soprano *et al.*, 2009).

Asimismo, el interés en el sector financiero ha crecido en las últimas dos décadas después de la aparición de algunos ejemplos destacados de pérdidas operativas debido a fraudes internos y externos, a daños a los activos físicos, a una disrupción de negocios y a fallas en los sistemas. La crisis crediticia desde 2008 en adelante ha aumentado definitivamente la atención de académicos, reguladores y profesionales sobre la reputación bancaria. A pesar de su importancia, pocos estudios proporcionan evidencia empírica sobre el riesgo reputacional en la industria financiera y la mayoría se centran en estimar el alcance de las pérdidas reputacionales como reacción del mercado al anuncio de pérdidas operativas basadas en una serie de eventos históricos (Soprano *et al.*, 2009). En general, se suele encontrar que los anuncios de pérdidas operativas generan daños estadísticamente significativos a la reputación.

Por ejemplo, Gillet *et al.* (2010) realizaron una evaluación de las reacciones de los mercados europeos y estadounidenses ante la comunicación de pérdidas operativas por empresas financieras entre 1990 y 2004. La finalidad de dicho estudio fue separar las pérdidas operativas del daño reputacional. Por otro lado, Fiordelisi *et al.* (2011), realizaron un estudio similar para estimar el impacto reputacional a partir del anuncio de pérdidas operacionales entre 1994 y 2008, suministrando nueva evidencia en el sector bancario a partir de los siguientes dos resultados principales,

- La probabilidad de daño reputacional aumenta a medida que aumentan las ganancias y el tamaño del banco.

- Un mayor nivel de capital invertido e intangibles reducen la probabilidad de daño reputacional.

Con respecto a la medición, Mitic (2018) emplea un modelo econométrico para medirlo en una organización, con la particularidad de que utiliza un indicador que cuantifica el sentimiento de corto plazo con variables financieras de la organización, tales como las ventas y los rendimientos. Asimismo, utiliza el VaR (*Value at Risk*) al 99,9% sobre la distribución de pérdidas para obtener una medida de riesgo más adecuada para focalizar sobre la cola de los eventos de pérdidas más significativas. Una de las principales conclusiones es que la reputación negativa tiene un impacto más significativo que la positiva sobre las variables medidas (ventas y ganancias). En el presente trabajo, se emplea un indicador de reputación de naturaleza continua para estimar el riesgo reputacional.

La medición de la exposición al riesgo reputacional no es un asunto sencillo (Soprano *et al.*, 2009), y resulta difícil separar un efecto puramente reputacional de otros tipos de riesgos. La identificación del componente específico de reputación, a diferencia de los riesgos crediticios, de mercado u operativos, requiere de procedimientos claros para determinar el capital en riesgo aceptable y evitar un sobrecargo de capital injustificado. Para esto existen algunas técnicas alternativas de medición de este riesgo porque no es un campo consolidado.

El Riesgo Reputacional es uno de los riesgos al que está expuesta una institución financiera en los mercados actuales, especialmente aquellas empresas que cotizan en bolsa, ya que la ocurrencia de un evento con un efecto significativo y generalizado inevitablemente tendrá un impacto en el precio de las acciones y traerá una gran presión a los directivos de la organización. Además, existe la posibilidad de que los eventos que le suceden a determinadas instituciones financieras puedan afectar a todo el sector, ya que el mercado puede asumir que toda la industria está expuesta a riesgos similares (Soprano *et al.*, 2009). A continuación, se enumeran los posibles efectos de un evento de riesgo reputacional,

- Una caída en el precio de las acciones de la empresa. Esto podría durar un corto período de tiempo o tener un efecto más prolongado y duradero, según el proceso de gestión del riesgo reputacional de la empresa, la capacidad para comunicarse con los mercados financieros y el regulador, y las decisiones tomadas para reducir o manejar estas situaciones. Una empresa también debe considerar que la solidez y la coherencia en la gestión del riesgo reputacional a lo largo del tiempo tendrá un efecto a largo plazo en su credibilidad por parte de los inversores y de los clientes, y su sostenibilidad a largo plazo impactará en la distribución de los productos y servicios ofrecidos, en la aprobación por parte del regulador, como así también en la evaluación de las agencias calificadoras de riesgos y finalmente en su rentabilidad.
- Para muchos eventos reputacionales, el efecto más inmediato, y a menudo a largo plazo, será la pérdida de negocios e ingresos, ya que hará que los clientes opten por proveedores de servicios alternativos. Además, habrá mayores costos para recuperar la confianza de los clientes y del mercado.
- Puede haber un impacto relevante en la imagen y en la marca de la empresa, lo que requiere una revisión y una campaña para cambiar y/o modificar, por ejemplo, el logotipo, la estrategia publicitaria y/o la política de comunicación del mercado. Incluso podría obligar a la empresa a salir de ciertos mercados.

- Puede haber eventos que afecten la confianza de los empleados en la estrategia o en la confiabilidad de la empresa, causando desmotivación, alta rotación del personal y dificultades para atraer capital humano.
- Un incremento de la auditoría por parte del mercado (con autoridades reguladoras y agencias de calificación) podría generar mayores costos para el personal y para los sistemas dedicados a los controles internos.

En los últimos años, las consultoras han comenzado a alertar sobre la importancia de gestionar el riesgo reputacional. En este sentido, KPMG ha realizado varios avances en materia de concientización sobre este riesgo. En la siguiente imagen se sintetiza en forma precisa la pretensión de los reguladores acerca de este riesgo en un **informe** de la consultora internacional KPMG.



Figura 2. KPMG - Informe 2019

KPMG (2019) recomienda a las Entidades Financieras incluir al Riesgo Reputacional dentro de su política integral de gestión de riesgos mediante i) la inclusión en el gobierno, ii) la generación de reportes, iii) la utilización de herramientas para medirlo y dimensionarlo y iv) la integración en la gestión de la organización. Esto se resume en forma clara en la siguiente imagen extraída del informe precedente.



Figura 3. KPMG - Informe 2019

Además, la consultora señala diez claves que debería considerar cualquier organización, y en particular una compañía financiera para gestionar el riesgo reputacional (KPMG, 2019):

1. **Abordar** el riesgo reputacional desde las dos perspectivas: riesgos y reputación. Los responsables de los riesgos deben validar y promover la metodología.
2. **Incluir** todas las fuentes en el momento de identificar los eventos de riesgo, entre otros, los mapas de riesgos de la organización o las reclamaciones recibidas por la organización.
3. **Evaluar** los eventos de riesgo contemplando tanto la perspectiva interna (direcciones involucradas) como la externa (grupos de interés).
4. **Priorizar** los eventos, clasificados por severidad, para gestionar de manera eficaz los que presenten una mayor amenaza.
5. **Cuantificar** el impacto económico para facilitar la toma de decisiones de los órganos de gobierno y cumplir con las expectativas de los supervisores.
6. **Hacer** un seguimiento periódico de la gestión del riesgo reputacional.
7. **Supervisar** los indicadores de desempeño, la sensibilidad social y la gestión de los grupos de interés para poder anticiparse a la materialización de posibles eventos.
8. **Disponer** de una política de riesgo reputacional que defina, principalmente, el modelo de gobierno, la integración en la gestión, el reporting y el entorno tecnológico.
9. **Automatizar** la gestión del riesgo reputacional, permitiendo una visión integrada del modelo, mayor trazabilidad, menor carga de trabajo y menor riesgo operacional.

10. **Contar** con herramientas de análisis y manuales de gestión de crisis que permitan una respuesta inmediata en caso de que se materialice un evento de riesgo reputacional.

En el presente trabajo se incluyen a todas las fuentes del riesgo reputacional, se cuantifica el impacto económico mediante el cálculo del Capital Económico y se propone una alternativa de monitoreo mediante la automatización del indicador de riesgo reputacional.

Por tal motivo, es importante que las empresas en general y los Bancos en particular analicen y evalúen claramente el impacto financiero de los riesgos reputacionales, proporcionando estimaciones de eventos extremos o críticos a la gerencia, de manera que se pueda identificar una estrategia adecuada de gestión de este riesgo.

1.3. Acerca de la Medición del Riesgo Reputacional

Según Soprano *et al.* (2009) consideran que modelar este riesgo como si fuese uno Operacional no es necesariamente la mejor forma, ya que los efectos son diferentes y provocan algo más que una pérdida inmediata. El impacto de un deterioro en la reputación de la empresa trasciende la cuenta de Pérdidas y Ganancias y representa una disminución de los Activos Intangibles. En esta primera parte del trabajo se presentarán algunas de las técnicas existentes de medición. En el capítulo 3 se propondrá una metodología propia para cuantificarlo.

La literatura presenta tres métodos diferentes para medir el riesgo reputacional,

1. Un modelo impulsado por la volatilidad del precio de las acciones.
2. Un enfoque basado en escenarios, similar al modelado cualitativo de riesgo operativo.
3. Mediante técnicas de tarjetas de puntuación.

El primer método mide el riesgo reputacional en función de la volatilidad del precio de las acciones. Si se supone la hipótesis de eficiencia del mercado, la cotización bursátil de una empresa refleja el valor presente de todos sus beneficios futuros. Así, cualquier información nueva que afecte a la capacidad de una empresa de generar beneficios en el futuro se verá reflejada inmediatamente en su cotización. Esto por supuesto incluye el impacto que eventos de tipo reputacional pueden tener en las ventas futuras de una empresa. Así, una manera de explorar el riesgo reputacional de una empresa puede ser evaluar empíricamente cómo eventos reputacionales negativos afectan a su cotización.

En su versión más sencilla, un modelo APT relaciona el retorno (rendimiento) R_t de un activo al momento t con el retorno de un portafolio de mercado $R_{Mkt,t}$ ⁴. Esta última variable puede representarse por el rendimiento de un índice accionario que refleja la evolución del mercado en el cual el activo bajo análisis se encuentra circunscripto. En términos formales,

⁴Es el acrónimo en inglés de *Arbitrage Pricing Theory* que emplea un modelo de valuación de precios de activos con múltiples factores basado en la idea de que los rendimientos de un activo se pueden predecir utilizando la relación lineal entre el rendimiento esperado del activo y una serie de variables macroeconómicas que capturan el riesgo sistemático.

$$R_t = \alpha + \beta_1 R_{Mkt,t} + \epsilon_t \quad (1)$$

Donde,

- α es la parte de R_t que no se puede explicar por $R_{Mkt,t}$
- β_1 mide la sensibilidad de R_t a las variaciones de $R_{Mkt,t}$
- ϵ_t es el término de error de la acción en el tiempo t

Para identificar el factor de riesgo reputacional, Soprano *et al.* (2009) proponen incluir la variable $R_{Rep,t}$, que reflejará la ocurrencia de eventos reputacionales negativos. Esta variable es dicotómica y toma valor igual a uno cuando ocurre un evento de reputación en el tiempo t , e igual a cero en caso contrario. Para construirla, la empresa tendrá que revisar todos aquellos sucesos pasados que afectaron su reputación, e identificar aquellos que se considera que tuvieron un impacto negativo.

El rendimiento R_t del precio genérico de las acciones de una organización en el tiempo t se describe por lo tanto mediante la siguiente ecuación,

$$R_t = \alpha + \beta_1 R_{Mkt,t} + \beta_2 R_{Bank,t} + \beta_3 R_{Rep,t} + \epsilon_t \quad (2)$$

donde cada elemento del vector β_j ($j = 1, 2, 3$) explica la sensibilidad de R_t a variaciones en los factores correspondientes y $R_{Bank,t}$ es una variable de control que refleja el retorno del sector bancario (o financiero).

Para calcular el Capital Económico (en Riesgo) por riesgo reputacional (VaR_{Rep}), Soprano *et al.* (2009) proponen determinarlo mediante la siguiente ecuación,

$$VaR_{Rep} = V_t * \delta * \beta_+ \quad (3)$$

Es decir, el capital en riesgo surge del producto de tres factores. El primer término V_t refleja el valor de mercado de la empresa (o la posición en riesgo) que surge del producto entre el precio de cierre ajustado de la acción P_t y la cantidad de acciones en circulación Q_t . Formalmente,

$$V_t = P_t * Q_t \quad (4)$$

El segundo término δ es la estimación puntual de un modelo de regresión *probit* o *logit*, donde se considera como variable respuesta (o dependiente) a $R_{Rep,t}$ que considera la ocurrencia de eventos reputacionales negativos.

$$\delta = \beta' x_i \quad (5)$$

Una desventaja importante de la metodología propuesta por Soprano *et al.* (2009) es que no hacen ninguna propuesta sobre cuáles deberían ser las variables independientes en ese modelo. Tener presente que la variable respuesta es binaria y que solamente puede asumir dos valores, cero o uno. Por lo tanto, si la estimación fuese igual a cero, no se estima un capital en riesgo.

Por último, el tercer término β_+ refleja el error de estimación del regresor $R_{Rep,t}$ en la especificación (3), que se determina como la multiplicación entre el factor de confianza y el desvío del estimador. Formalmente,

$$\beta_+ = t_{1-\frac{\alpha}{2}} * \hat{\sigma}(\hat{\beta}_3 R_{Rep,t}) \quad (6)$$

Donde $\hat{\sigma}$ representa el desvío estándar muestral del estimador del parámetro poblacional vinculado a la variable reputación. Esta metodología resulta ciertamente poco convencional, pues calcula el *VaR* a partir de los valores extremos del estimador, y no de la variable independiente.

Alternativamente, en el capítulo 3 del presente trabajo se presenta una metodología que observará los eventos reputacionales a partir del sentimiento en *Twitter*, construyendo una variable continua, que no sólo permitirá estimar los coeficientes de la regresión (2), sino que permitirá calcular el *VaR* atendiendo a los eventos más extremos de su distribución.

El segundo método mide el riesgo reputacional mediante la utilización de distintos escenarios. Un escenario se define como un evento hipotético que aún no ha ocurrido pero que podría afectar a la Entidad Financiera. Cada evaluación de escenario debe realizarse mediante extensas entrevistas a los directivos. Los pasos habituales que se utilizan para un análisis completo de escenario son,

- identificación de riesgos individuales y procesos críticos
- definición de escenario
- evaluación de escenarios
- evaluación final

Dado que no se puede realizar el análisis de escenarios para todos los procesos, se debe elaborar una lista y darle prioridad a aquellos que posean una mayor exposición a riesgo.

Teniendo en cuenta que durante la entrevista se efectúa una estimación cuantitativa, es muy importante recopilar la siguiente información,

1. Verificar que existan controles en caso de que ocurra el evento del escenario.
2. Estimar la probabilidad de que tal evento pueda ocurrir en forma anual.
3. Determinar el monto de pérdida promedio que podría implicar el escenario.
4. Estimar el monto máximo de pérdida en el evento del escenario.
5. Verificar si existe cobertura de seguro.
6. Recopilar sugerencias de alternativas de mitigación de riesgos.

Esta información se utiliza para calcular el Capital Reputacional en Riesgo (Capital Económico) del escenario seleccionado⁵. Con lo recolectado durante la entrevista, podría implementarse un modelo actuarial para estimar las distribuciones de severidad y frecuencia. Los pasos para estimarlo son,

⁵Esta información como por ejemplo la frecuencia por año, el monto promedio de pérdida, el monto máximo de pérdida y la cobertura del seguro.

- Se utiliza la frecuencia promedio por año para estimar la distribución de frecuencias.
- El monto promedio y máximo de la pérdida se aplican para estimar la distribución de la severidad (daño).
- La distribución anual de pérdidas se obtiene mediante simulación Monte Carlo.
- La distribución se ajusta a la cobertura del seguro.
- La pérdida esperada se calcula como la media (o mediana) de la distribución anual de la pérdida.
- La pérdida inesperada se calcula como la diferencia entre los percentiles a un nivel de confianza fijo (por ejemplo, 99,9 %) y la pérdida esperada.

Una vez completado el análisis de escenarios, los resultados deben compartirse con las funciones involucradas, presentando posibles pérdidas y recomendaciones para la mitigación de riesgos.

Luego de que se enumeran las áreas expuestas al riesgo reputacional, se mapean y se derivan, por ejemplo, mediante cuestionarios las clasificaciones de riesgo tanto en términos de frecuencia como de severidad. Se pueden estimar las pérdidas por riesgo operativo, de mercado o crédito y luego ajustar los resultados utilizando estos factores multiplicadores de reputación (Bessis, 2002).

Una desventaja de esta metodología es que resulta muy complicado encontrar una asunción razonable de la correlación entre los distintos eventos. Asimismo, se hace difícil estimar la probabilidad de sucesos tan infrecuentes y a veces se tendrá que utilizar la probabilidad subjetiva percibida por dos o más entrevistados. Esto no resulta deseable, pues los individuos no son buenos estimando probabilidades (Andersen *et al.*, 2014).

Por otra parte, esta metodología requiere de una evaluación no solo de los efectos directos de un evento catastrófico, sino también de los indirectos. Para los propietarios de los procesos podría resultar complejo, aunque se crea que son los más capacitados para evaluar las consecuencias directas de un evento que afecta sus procesos. Por lo tanto, frente a estas cuestiones el modelo APT estima en forma directa el impacto en el valor total de la compañía lo que resulta ventajoso para la medición del riesgo reputacional.

El tercer y último método se basa en tarjetas de puntuación para medir el riesgo reputacional. En cuanto a otras funciones de evaluación de control, ya sean auditoría interna, riesgo operativo, control de nivel de servicio o encuestas de satisfacción al cliente, se descubrieron que los cuestionarios o tarjetas de puntuación, también podrían aplicarse a la medición del riesgo de reputación.

De hecho, las listas ponderadas de preguntas, que deben completar los gerentes, propietarios de procesos, clientes o especialistas comerciales, podrían resultar particularmente útiles e indicar el valor de los impactos multidimensionales de los eventos de tipo reputacional⁶. Las tarjetas de puntuación, de hecho, tienen algunas características que pueden resultar adecuadas para la evaluación de este riesgo,

⁶Por impactos multidimensionales, se entienden que afectarían los ingresos, las relaciones con los clientes, los procesos internos, la tecnología de la información o la capacitación del personal, la percepción interna de la empresa, el precio y la calificación de las acciones, la cobertura de los medios y la valoración regulatoria.

- Están construidas a medida, lo que las hace extremadamente flexibles y fáciles de adaptar al negocio, al proceso y a la organización de la empresa⁷.
- Permiten centrarse y ampliar temas o áreas específicas de exposición.
- Pueden estructurarse de forma sencilla y, por tanto, fáciles de comprender, lo que hace que sea relativamente sencillo tanto como para que las personas lo completen como para captar los resultados.
- Las áreas de debilidad y la intervención necesaria se identificarán fácilmente, ya que se mostrarán en los cuestionarios.
- La estructura - o modelado - de preguntas y pesos asignados se puede modificar muy fácilmente.

Frente a estas ventajas, probablemente el mayor desafío para todos los modelos cualitativos basados en juicios de expertos es que los resultados pueden ser cuestionables y puede resultar difícil garantizar la coherencia de la evaluación,

- Diferentes analistas o personal entrevistado podrían determinar respuestas y evaluaciones muy diferentes de la probabilidad o impactos de eventos o los resultados de una situación de tipo reputacional. Por un lado, esto complica la integración y la lectura de dicha información y, por otro lado, dificulta enormemente el desarrollo de un cuestionario que asegure resultados consistentes.
- Dado que los resultados dependen de opiniones y juicios individuales, estos pueden presentar problemas al comparar los resultados, mostrando a veces un aumento inesperado en el perfil de riesgo o, por el contrario, un comportamiento estable invariable.
- Sin cambiar la muestra entrevistada, se experimenta que las mismas personas dan las mismas respuestas a lo largo del tiempo, simplemente confirmando sus valoraciones anteriores.
- Cuanto más articulado sea el cuestionario, más tiempo llevará completarlo, lo que genera riesgos de enfoque limitado o dificultades para llenarlo. A veces es difícil y requiere mucho tiempo repetir el ejercicio y extenderlo a un gran número de personas.

Una desventaja importante de estos modelos para el propósito de este trabajo es que no permiten estimar un capital en riesgo al ser puramente cualitativo. Frente a esto, el modelo propuesto en el capítulo 3 presenta varias ventajas. En primer lugar, al considerar todos los riesgos reputacionales en un indicador, resuelve el problema de las correlaciones. En segundo lugar, al construir el indicador a partir de los comentarios de la red social *Twitter*, permite estimar la probabilidad de los sucesos negativos a partir de los datos, sin la necesidad de recurrir al criterio humano.

⁷Y posiblemente no existe una lista de verificación estándar de preguntas que las asociaciones de la industria o los reguladores recomienden usar para la valoración del riesgo de reputación.

1.4. En búsqueda de la Solvencia

El Comité de Supervisión Bancaria de Basilea es el organismo a nivel internacional que emite normas y recomendaciones para tratar de mejorar y armonizar las reglas aplicables a la banca y, en particular, sobre sus recursos propios mínimos exigibles.

Este Comité no debe considerarse como una autoridad de supervisión supranacional, sino más bien como un organismo que persigue proveer a sus miembros de estándares y guías de supervisión para mejorar sus prácticas, con la expectativa de que las autoridades de cada país tomen las medidas que faciliten su implementación para lograr una mejora en la posición internacional de sus propios sistemas nacionales.

Hasta el momento se realizaron tres acuerdos de Basilea. El primero fue en 1988 y solo se centraba en el Capital de Riesgo de Crédito y de Mercado (incorporado en 1997). Era un modelo simplista ya que no consideraba la diferente calidad crediticia de los clientes, los efectos de la correlación y diversificación y no tenía incentivos para la mejora de la medición y gestión del riesgo.

En Basilea II (2004) se complementan y amplían las normas iniciadas en Basilea I respecto a la exposición a los diferentes riesgos y a la suficiencia de los fondos propios y establece nuevas exigencias a la banca sobre transparencia y publicación periódica de la información⁸. Con este acuerdo se reconoce un nuevo riesgo: el Operacional. Adicionalmente, para Riesgo de Crédito se establece la posibilidad de utilizar ponderaciones específicas en función de la calidad crediticia de cada cliente, se permite el uso de calificaciones otorgadas por agencias de calificación externas y el uso de métodos de cálculo basados en modelos internos de calificación siempre que hayan sido aprobados por el supervisor. También se introduce un tratamiento del Riesgo de Mercado más sensible al perfil del riesgo de la cartera.

El énfasis se pone en la seguridad y estabilidad del sistema bancario (tanto nacional como internacional) a través de tres pilares. El primero tiene como objetivo garantizar la solvencia de las entidades exigiendo unos requerimientos de Capital mínimos para Riesgo de Crédito, Mercado y Operacional. El segundo fomenta que los Bancos cuenten con procesos y sistemas para evaluar la suficiencia de Capital total en función de su perfil de riesgo y con una estrategia de mantenimiento del mismo por encima de los coeficientes mínimos. Además, se exige que la alta dirección del Banco se involucre activamente en el control de riesgos y en la planificación futura de las necesidades de Capital. Y el último establece normas de transparencia y exige la publicación periódica de información acerca de la exposición de los Bancos a los diferentes riesgos y a la suficiencia de sus fondos propios.

Al concepto de Capital se le hace alusión en los Pilares I y II. En el primero se hace referencia al Capital Regulatorio y en el otro, al Capital Económico. A continuación, se explicarán ambos conceptos, sin embargo en este trabajo el foco estará en el último.

El tercer acuerdo (Basilea III) fue en el año 2010, como respuesta a la crisis Subprime (2008) con el objetivo de fortalecer la capacidad de resistencia de la banca, mejorando la calidad y la cantidad del capital de las Entidades. También busca reducir la prociclicidad del Sistema Financiero mediante un mecanismo amortiguador anticíclico. Además, se agregan conceptos de liquidez y apalancamiento.

⁸Basel Committee on Banking Supervision, International Convergence of Capital Measurement and Capital Standards. A Revised Framework, Basel, 2004.

El Banco Central de la República Argentina (BCRA) está alineado a las recomendaciones que promulga el Comité. El Capital Regulatorio es la cantidad de recursos propios que el regulador considera que una Entidad Financiera debe reservar como mínimo para garantizar la solvencia frente a los diferentes riesgos (Crédito, Mercado y Operacional)⁹. Como se mencionó, en la actualidad no se reserva Capital Regulatorio por Riesgo Reputacional.

En la Comunicación “A” 5398, el BCRA determina que es de fundamental importancia que las Entidades Financieras cuenten con un proceso interno, integrado y global, para evaluar la suficiencia de su capital económico en función de su perfil de riesgo (*Internal Capital Adequacy Assessment Process - ICAAP*) y con una estrategia para mantener sus niveles de capital a lo largo del tiempo.¹⁰

El cálculo del Capital Económico permite determinar los recursos que debe mantener una entidad para poder cubrir las pérdidas inesperadas. Al definirlo debe hacerse referencia a los conceptos de pérdida esperada, la que constituye el valor esperado de la distribución de pérdidas. La siguiente figura exhibe la determinación del Capital Económico como la diferencia entre ambas medidas. La pérdida esperada corresponde a un promedio de pérdida, que se debe tomar en cuenta como gasto inherente del negocio de crédito, ya sea por medio de las provisiones o como parte de la tasa de interés, de modo análogo a una prima de riesgo en el área de seguros. La pérdida inesperada, por otro lado, corresponde a fluctuaciones o picos sobre el saldo promedio y no se puede predecir con certeza. De hecho, siempre existe la posibilidad de experimentar un desvío mayor que la máxima pérdida histórica. Por lo tanto, esas fluctuaciones desfavorables deben ser cubiertas con capital adicional a las provisiones con el efecto de absorber pérdidas potenciales imprevistas que puedan poner en riesgo el Capital de la Entidad.

La distribución de pérdidas habitualmente no es simétrica, sino que cuenta con un sesgo positivo (hacia la derecha), lo que indica que existe cierta probabilidad positiva de experimentar pérdidas extremas. Se define como el percentil $k\%$ al valor monetario de la distribución de pérdidas, tal que el $k\%$ de los posibles valores de pérdida estén por debajo de ese valor. El percentil de la distribución de pérdidas es lo que comúnmente se denomina “Valor en Riesgo” (*VaR*), y es uno de los modos más convenientes de especificar los valores de pérdidas que se encuentran en el extremo de la distribución. Por lo tanto, el *VaR* a cierto nivel de confianza consiste en especificar el percentil de la distribución estadística de pérdida que deseamos cubrir.

Para cubrir el riesgo de crédito, Basilea recomienda utilizar un nivel de confianza del 99,9%, lo que significa cubrir la pérdida esperada con provisiones, y la diferencia entre el percentil de pérdida del 99,9% (*VaR*) y la pérdida esperada con un requerimiento de capital, tal como se detalla a continuación.¹¹

⁹La normativa referida al capital regulatorio es la Comunicación “A” 5831.

¹⁰Proceso interno de evaluación de la adecuación de capital. La normativa referida es la Comunicación “A” 5398, punto 1.3.

¹¹Basel Committee on Banking Supervision, International Convergence of Capital Measurement and Capital Standards. A Revised Framework, Basel, 2004.

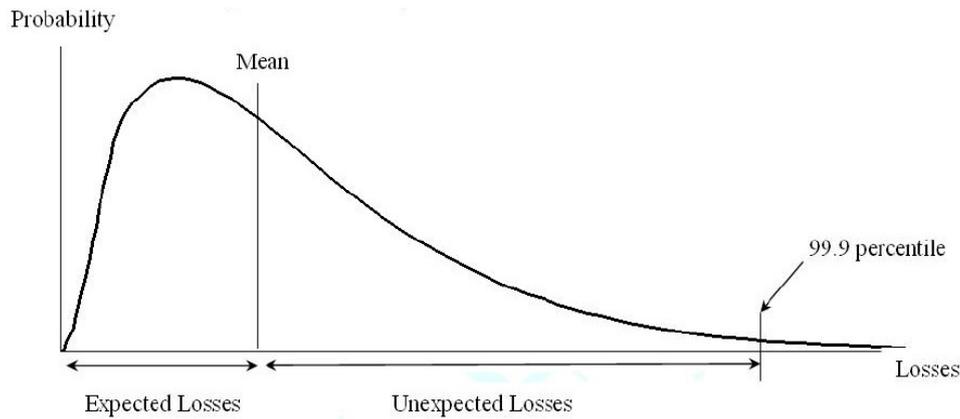


Figura 4. Distribución de Probabilidad de Pérdidas Potenciales

Por lo tanto, la fórmula de cálculo sería la siguiente,

$$\text{Capital Económico} = VaR_{k\%} - \text{Pérdida esperada} \quad (7)$$

El Capital Económico tiene la ventaja de ser una medida más adecuada para medir el riesgo que la utilizada por los regímenes regulatorios vigentes. Esta fórmula establece el capital que una entidad financiera debería reservar para cubrir, no sólo las pérdidas inesperadas originadas por las exposiciones a un determinado riesgo al que podría estar expuesta la compañía. Se determina como la diferencia entre el Valor a Riesgo a un determinado nivel de confianza, indicando la peor pérdida en condiciones normales de mercado y la pérdida esperada que se determina como el capital mínimo reservado por normativa para los riesgos de mercado, operacional y crédito. Cada área de Gestión de Riesgos de las Entidades Financieras deben desarrollar sus modelos internos para estimar el Capital Económico por cada uno de los riesgos a los que está expuesta según lo establecido por la normativa local y las recomendaciones internacionales de Basilea¹².

Los modelos internos de Capital Económico se han erigido como una importante herramienta de gestión para la medición, el control y la gestión integrada de los riesgos, la gestión estratégica del negocio y de los recursos disponibles, la valoración del retorno ajustado al riesgo de los negocios (RAROC) y el establecimiento de márgenes de precios de referencia.

El Comité de Supervisión Bancaria de Basilea planteó como principal preocupación la necesidad de formar un capital suficiente para absorber las pérdidas inesperadas que enfrentan las Compañías Financieras. Por lo cual, es preciso tener conocimiento del nivel de riesgo que asumen para proteger la solvencia del sistema financiero y minimizar el riesgo de quiebra. Asimismo, creen que las ventajas de un sistema en el cual el Capital está más cercano al riesgo asumido superan claramente sus costos, obteniendo como resultado que el sistema bancario sea más seguro y más eficaz.

En este trabajo, se realizará una estimación del Capital Económico por Riesgo Reputacional mediante la utilización de técnicas de *Text Mining*, lo cual representa un enfoque novedoso.

¹²Comunicación “A” 5398 - Lineamientos para la Gestión de Riesgos en las Entidades Financieras

Resumen de Capítulo

En el presente capítulo se define al riesgo reputacional como un riesgo exótico en las organizaciones en general y en particular en las Entidades Financieras. Para eso se deja evidencia sobre la falta de una visión unificada para cuantificarlo. Esta tesis brinda una herramienta para medir el riesgo reputacional y calcular el Capital Económico que deben reservar los Bancos para hacer frente ante eventuales efectos adversos originados por una percepción negativa mediante la utilización de un modelo econométrico.

En el siguiente capítulo profundizará en la utilización de *Text Mining* para medir el riesgo reputacional a partir de datos extraídos de la red social *Twitter*.

Esta página se deja en blanco intencionalmente

"...Hay secretos en el fondo del mar,
personas que me quiero llevar,
aromas que no voy a olvidar, silencios
que prefiero callar, mientras vos
jugás..."

Fito Páez, *Brillante sobre el mic*

Capítulo 2: Text Mining aplicado a la medición del Riesgo Reputacional

En el capítulo anterior se le brindó una herramienta importante a un gerente de riesgos que le permite alocar capital para el riesgo reputacional. En este capítulo se profundiza en la utilización de *Text Mining*, en particular del Análisis de Sentimiento para medir el Riesgo Reputacional de Entidades Financieras.

Para cumplimentar con dicho objetivo, se introducirá el concepto de *Big Data*, luego se desarrollará el concepto de *Text Mining* para vincularlo con la aplicación en el Riesgo Reputacional. Posteriormente, se deriva el Análisis de Sentimiento como una herramienta que permite convertir información textual en cuantitativa para su posterior análisis y procesamiento. Asimismo, se presenta la metodología de cálculo del indicador de reputación. En el capítulo se concluye mediante la exhibición de varios resultados obtenidos en ciertos trabajos académicos sobre la utilización de estas técnicas con los movimientos de los precios de mercado.

2.1. Sobre el análisis de sentimiento

El término *Big Data* se presenta como una tecnología disruptiva que está revolucionando la manera en que funciona el mundo, con la capacidad de impactar de forma estratégica en toda la sociedad (Munafo, 2019), y de generar transformaciones productivas, económicas y sociales de gran envergadura (Schmarzo, 2013). A diferencia de otras revoluciones tecnológicas, el impulso se focaliza en la transformación, el análisis, el uso y el almacenamiento de grandes cantidades de información de manera automatizada (McAfee *et al.*, 2012). También se lo denomina macrodatos o datos masivos como un concepto que se refiere al almacenamiento de grandes cantidades de datos y a los procedimientos utilizados para encontrar patrones repetitivos dentro de estos (Schmarzo, 2013).

Aunque las definiciones no sean coincidentes entre sí, la mayoría concuerda en la capacidad de encontrar patrones que permiten agregar valor a la organización, con la finalidad de tomar mejores decisiones.

El *Text Mining* es el conjunto de técnicas y tecnologías que se utilizan para explorar grandes cantidades de texto, de manera automática o semiautomática, y descubrir patrones repetitivos, tendencias o reglas que explican el comportamiento del texto¹³. Asimismo, según Kwartler (2017) lo define como “*is the process of distilling actionable insights from text.*”, es decir como aquel proceso de obtener información procesable del

¹³*Text Mining* se denomina en castellano Minería de Textos.

texto que puede contribuir a varias unidades dentro de una organización. El *Text Mining* comprende tres actividades fundamentales,

- Recuperación de la información: Consiste en seleccionar los textos pertinentes
- Extracción de la información incluida en esos textos mediante el procesamiento del lenguaje natural: Hechos, acontecimientos, datos clave, relaciones entre ellos, etc.
- Minería de datos para encontrar asociaciones entre los datos clave previamente extraídos de entre los textos.

Una definición formal indica que *Sentiment Analysis* es el campo de estudio que analiza las opiniones, sentimientos, valoraciones, actitudes y emociones de las personas hacia las entidades y sus atributos expresados en texto escrito (Liu, 2015)¹⁴. Las entidades analizadas pueden ser muy variadas, como por ejemplo productos, servicios, organizaciones, individuos, eventos, problemas o temas en general.

Según Liu (2015) a veces se confunde un sentimiento de una opinión. El primero se define como una actitud, pensamiento o juicio impulsado por un sentimiento, mientras que la segunda como una visión, juicio o valoración formada en la mente sobre un asunto en particular. La diferencia es bastante sutil y cada uno contiene algunos elementos del otro. Las definiciones indican que una opinión es más la visión concreta de una persona sobre algo, mientras que un sentimiento es la experiencia subjetiva de las emociones.

Por ejemplo, la oración “Me preocupa el estado actual de la economía” expresa un sentimiento, mientras que “Creo que la economía no va bien” expresa una opinión. En una conversación, si alguien dice la primera oración, podemos responder diciendo, “Comparto su preocupación”, pero para la segunda oración, normalmente diríamos, “Estoy de acuerdo/en desacuerdo con usted”. Sin embargo, los significados subyacentes de las dos oraciones están relacionados porque es probable que el sentimiento descrito en la primera sea causado por la opinión de la segunda. A la inversa, también se puede decir que la primera oración implica una opinión negativa sobre la economía, que es lo que dice la segunda oración.

En este trabajo se utiliza el término sentimiento para referirse al sentimiento subyacente positivo o negativo implícito en la opinión.

El análisis de sentimientos se centra por tanto en opiniones que expresan o implican sentimientos positivos o negativos, y al analizarlos también se debe considerar aquellas expresiones sin ningún sentimiento implícito, es decir las denominadas neutrales. Las oraciones que expresan opiniones o sentimientos suelen ser oraciones subjetivas en contraposición a las objetivas, que por ejemplo enuncian hechos, mientras que las opiniones y los sentimientos son inherentemente subjetivos. Sin embargo, las oraciones objetivas también pueden implicar sentimientos positivos o negativos de sus autores, porque pueden describir hechos deseables o indeseables. Por ejemplo, se sabe que “compré el auto ayer y se rompió hoy” y “después de dormir en el colchón durante un mes, se formó un hueco en el medio” describen dos hechos indeseables, y se puede inferir con

¹⁴Existen varios nombres relacionados con tareas levemente diferentes, como por ejemplo, análisis de sentimientos, minería de opiniones, análisis de opiniones, extracción de opiniones, minería de sentimientos, análisis de subjetividad, análisis de afectos, análisis de emociones y minería de opiniones. Todas se encuentran consideradas en el análisis de sentimientos.

seguridad que los autores de las oraciones tienen sentimientos negativos sobre el coche y el colchón. El análisis de sentimientos también estudia tales oraciones objetivas.

2.2. Acerca de la medición del Sentimiento y el modelo de Polarización

En la siguiente imagen se exhibe el diagrama de flujo con la modelización empleada en este trabajo, donde se mencionan cada uno de los pasos empleados para cumplimentar con el objetivo planteado.

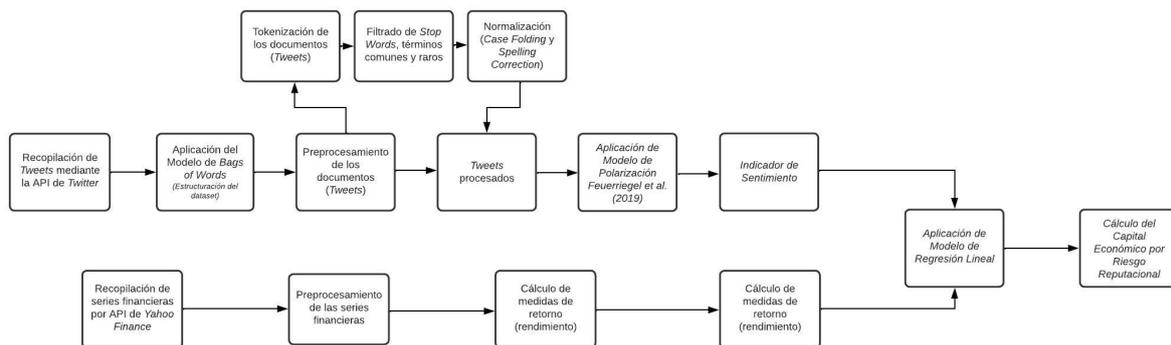


Figura 5. Diagrama de Flujo del Modelo

Tal como se exhibe en el diagrama, luego del proceso de recopilación de los *Tweets* se emplea el modelo de *Bag-of-Words* para procesar los documentos y representarlos ignorando el orden las palabras. Es importante destacar que en este modelo, las características que se desprenden de los documentos procesados se basan en las palabras que contienen¹⁵. Por tanto, cada documento está representado por una bolsa (multiset) de palabras. Las palabras que aparecen en el documento se encuentran presentes en este conjunto y aquellas que no lo están, no forman parte del mismo. Es importante aclarar que estos conjuntos se pueden convertir en vectores (Zizka *et al.*, 2020)¹⁶. A los fines del presente trabajo, se considera a cada *tweet* como un documento. En la siguiente figura se exhibe este enfoque,

¹⁵A las características también se las denominan atributos o *features* en inglés. En el presente trabajo se utilizará cada *tweet* recolectado como un documento utilizado para su procesamiento.

¹⁶Son una representación de datos más común en el aprendizaje automático.

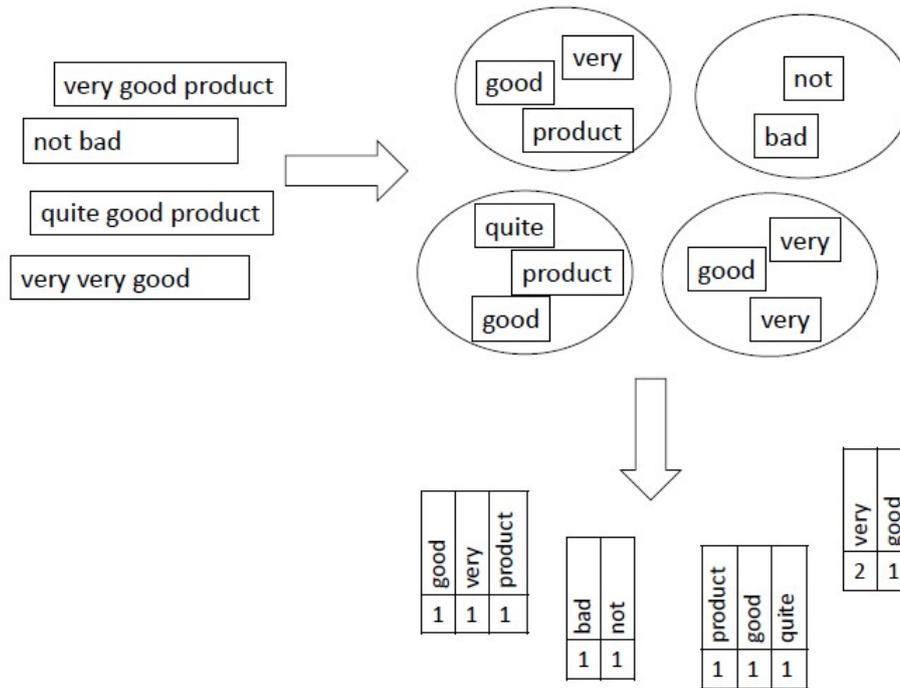


Figura 6. Modelo de *Bags-of-Words*

En el gráfico anterior se puede observar que diferentes documentos están representados por conjuntos de tamaños diferentes (que contienen diferentes números de elementos). Los vectores creados a partir de los conjuntos pueden tener tamaños diferentes. Y además se observa que los componentes vectoriales (encabezados de tabla) tienen diferentes significados. Por ejemplo, la primera componente del primer vector representa la presencia de la palabra *good* mientras que la primera del segundo vector representa la presencia de la palabra *bad*. Esta situación no es un estado deseado cuando se pretende, por ejemplo, comparar estos dos documentos. Tienen el mismo valor que la primera característica, pero el significado semántico es opuesto.

Estas consideraciones llevan a la conclusión de que, en la representación final de los documentos individuales a procesar, todos los documentos deben tener el mismo número de características y hay que tener la capacidad de identificar claramente cada característica en toda la colección de documentos. Por lo tanto, el conjunto de atributos para la colección debe acomodar todas las características posibles y todos los documentos compartirán las mismas características. Por supuesto, algunos de los documentos pueden no estar caracterizados por algunas características (es decir, su valor es 0), como en la imagen anterior. Por lo tanto, todos los documentos de una colección están representados por vectores del mismo tipo y se pueden organizar en una tabla conocida como *term-document matrix* (*tdm*)¹⁷.

La aplicación de este modelo en la creación del *tdm* tiene algunas desventajas. Primero, existe una gran cantidad de posibles características que pueden derivarse de los textos. Al considerar las palabras como las características más naturales, incluso una colección muy pequeña de documentos puede contener muchas palabras únicas.

¹⁷Dicha matriz tiene un formato altamente estructurado que es adecuado para muchos algoritmos de aprendizaje automático.

A pesar de los numerosos problemas y limitaciones de este modelo, ofrece resultados sorprendentemente buenos en muchas tareas y es adecuado para el problema propuesto en este trabajo. Una de las ventajas de este modelo radica en la facilidad de utilización y en la eficiencia computacional. Además, en ciertas situaciones este enfoque es apropiado, como por ejemplo cuando hay menos características, donde no hay correlación entre ellas o cuando no se desea tenerlas en cuenta (Zizka *et al.*, 2020).

Las palabras son las características más naturales y sencillas que pueden derivarse de los textos. Es una secuencia de símbolos, típicamente letras, que se usa en las oraciones de un idioma dado y que tienen una semántica, unidad gramatical y fonológica. Por lo general, está separada de otras palabras por espacios en la escritura¹⁸. Por lo tanto, la ausencia de una secuencia particular de caracteres en un diccionario no necesariamente significa que no es una palabra adecuada. Por lo general, depende de la tarea a realizar. Una de las desventajas de emplear palabras es que, sin un contexto, el posible significado de una palabra puede no ser obvio.

A los fines de la presente tesis se utiliza como unidad lingüística a la palabra de cada comentario en *Twitter* por ser el atributo más adecuado para analizar el sentimiento de los comentarios de los usuarios de la red social, dado que por ejemplo un *tweet* puede poseer varias palabras con connotación negativa y por lo tanto, debería clasificarse con un mayor grado de negatividad.

Con respecto al procesamiento del texto, existen distintas herramientas, tales como *tokenización* y detección de oraciones. La primera es el proceso de dividir un documento en fragmentos de texto conocidos como *tokens*, los cuales representan a menudo a las palabras contenidas en el texto. En la mayoría de los idiomas europeos donde las palabras están delimitadas por espacios, la tarea parece ser bastante simple para dividir el texto en los lugares donde hay espacios en blanco.

En algunas tareas, puede resultar beneficioso dividir un documento en unidades más grandes, como oraciones. Parece que encontrar símbolos que normalmente terminan las oraciones, como por ejemplo: “.”, “?”,”!” debería ser suficiente. Sin embargo, estos símbolos también pueden ser parte de otras expresiones (por ejemplo, un punto puede ser parte de una abreviatura o un número). También se puede omitir un símbolo de terminación, como en el caso de que una oración termine con una abreviatura.

En el presente trabajo se utilizan ambos procesamientos de los datos para emplear como unidad lingüística a la palabra, por lo tanto se procede a realizar una *tokenización* de cada comentario extraído de la red social. Con respecto a la detección de oraciones, se consideran los símbolos previamente descritos para indicarle al algoritmo la finalización de una oración.

Ahora bien, algunas de las palabras que son demasiado comunes no suelen contribuir (o contribuyen sólo de forma insignificante) al logro de un determinado objetivo (Zizka *et al.*, 2020). Las palabras que no son importantes para una determinada tarea se conocen como *Stop Words* y a veces son las más frecuentes en un determinado idioma¹⁹. El largo de la lista de palabras vacías, en comparación con el tamaño del diccionario, que puede ser de varios cientos de miles, es insignificante. Por lo tanto, el número de funciones no disminuye significativamente cuando se eliminan dichas palabras.

¹⁸Así es como entienden las palabras la mayoría de los idiomas occidentales.

¹⁹Las **Stop Words** son palabras vacías, como por ejemplo preposiciones de artículos gramaticales, conjunciones, verbos auxiliares, pronombres, artículos.

Por lo tanto, es posible crear una lista de palabras que se excluyen del análisis, teniendo en consideración aquellas que sean más frecuentes y por lo tanto más importantes. Sin embargo, debe considerarse cada palabra en el contexto en el cual se están analizando los datos. Por ejemplo, en este trabajo la palabra “Banco” será demasiado frecuente y no será útil para determinar si un comentario puede considerarse neutral, positivo o negativo respectivamente.

A los fines del presente trabajo, se incorpora un listado de palabras vacías y raras muy utilizadas en el idioma castellano y vinculadas al contexto del análisis financiero²⁰. Esta lista de palabras son filtradas del *dtm* y no se consideran al momento de extraer información de los datos por no aportar valor al análisis realizado. Es importante dejar en claro que esta técnica tiene la desventaja de que la lista utilizada no sea exhaustiva de la totalidad de palabras vacías y raras vinculadas al contexto de análisis. Por lo tanto, se presenta como una limitación del presente trabajo.

En este trabajo se aplicó la técnica de preprocesamiento Normalización de los *tokens*, la cual consiste en colocar el texto en igualdad de condiciones (Turney *et al.*, 2010). Por ejemplo, están escritos en un tipo diferente de letra, sin embargo tienen el mismo significado en la misma forma. Ahora bien, existen distintos tipos de transformaciones, como por ejemplo *Case Folding*, *Spelling Correction*, entre otras.

La primera es el proceso de convertir todos los caracteres de una palabra a minúsculas o mayúsculas, por lo general se convierten a minúsculas. Cuando dos o más palabras con el mismo significado se escriben con mayúsculas y minúsculas diferentes, el número de palabras únicas aumenta innecesariamente y, por lo tanto, es favorable convertirlas a una misma forma. Las razones por las que la misma palabra se escribe con mayúsculas y minúsculas diferentes pueden incluir la posición al principio de una oración, la presencia en un título, la intención de enfatizar una palabra, etc.

Por otro lado, algunas palabras requieren mayúsculas y, en estas situaciones, modificarlas cambia el significado de las mismas. Además, a veces en internet, las mayúsculas pueden representar gritos y por lo tanto, convirtiéndose en una manera de expresar entonación en un medio escrito, vinculándose con el enfado, la agitación o el nerviosismo.

Por último, los errores de tipeo son problemas comunes en cualquier texto escrito, y traen ruido a los datos, por lo tanto puede complicar la consecución de resultados satisfactorios en algunas tareas. La corrección ortográfica (*Spelling Correction*) es una solución a este problema. Por lo general, este proceso implica la detección de un error, la generación de correcciones candidatas y la clasificación de las mismas (Kukich, 1992).

En el presente trabajo se normalizan los *tokens* mediante las transformaciones descriptas. Es decir, se convierten todas las palabras de cada *tweet* a minúsculas y en caso de errores de tipeo, el algoritmo empleado en este trabajo realiza correcciones ortográficas.

Luego de la etapa previa de procesamiento de los datos, es decir que se tiene en claro que *features* se utilizarán para describir el documento procesado, es necesario asignarles un valor (o peso) a cada *tweet*. Por lo tanto, a cada atributo de cada documento se le asigna un peso (o ponderación) para representar su importancia. El valor se calcula en función de la cantidad de apariciones de las palabras en cada *tweet* y en toda la colección.

La ponderación w_{ij} de cada término i en el *tweet* j se conforma de tres componentes,

²⁰El algoritmo empleado en este trabajo tiene incorporado un diccionario financiero para realizar el análisis de sentimiento.

1. Peso Local: lw_{ij} que representa la frecuencia en cada *tweet*.
2. Peso Global: gw_i que exhibe la capacidad discriminativa del término, basado en la distribución del mismo en toda la colección de *tweets*.
3. Factor de Normalización: n_j que corrige el impacto del largo de los diferentes *tweets*.

La ponderación w_{ij} se calcula de la forma siguiente,

$$w_{ij} = \frac{lw_{ij} * gw_i}{n_j} \quad (8)$$

El factor lw_{ij} cuantifica la importancia de un término en un solo documento, sin considerar los otros documentos. Por lo tanto, se basa en la frecuencia de aparición del término en el *tweet* específico.

Existen distintas formas funcionales de este peso, por ejemplo, en algunas situaciones, solamente es importante si una palabra está presente o ausente en un *tweet*. En esta situación, lw_{ij} será binario, y asumirá el valor 0 cuando un *tweet* no contiene un término, y el valor 1 en caso contrario independiente de su frecuencia.

El método de ponderación de términos locales más utilizado es la cantidad de veces que aparece un término (*Term Frequency*). Es decir, su frecuencia absoluta simple. En el presente trabajo se utiliza esta forma funcional del peso local (*Term Frequency*) por representar una ponderación de fácil entendimiento y que brinda el peso a cada término según las veces que aparece. En el apéndice 2 se exhiben detalles sobre las distintas formas funcionales de cálculo del peso local.

Con respecto a las ponderaciones globales, el objetivo es disminuir la importancia de los términos que son demasiado comunes. Por ejemplo, algunas palabras en inglés aparecen en casi todos los textos en dicho idioma. Por lo tanto, la finalidad es disminuir el peso de este tipo de palabras y dar uno mayor a las que no aparecen con tanta frecuencia para que tengan un mayor poder discriminativo. Esto significa que se debe considerar la distribución de palabras en toda la colección de documentos. Ahora bien, hay ciertos términos que a pesar de aparecer con bastante frecuencia poseen la importancia asignada por el peso local que en este trabajo se eligió ponderar por las veces que se presentan en el documento.

La *Inverse Document Frequency* (IDF) es el peso global más popular y se determina como el logaritmo de la probabilidad inversa de que un término aparezca en un documento aleatorio. Cuando un término está contenido en cada documento, la probabilidad de que aparezca es 1 y el logaritmo es igual a 0. Esto significa que este término no tiene impacto en ningún cálculo posterior. A los fines del presente trabajo se utiliza esta forma funcional de la ponderación global por ser la más utilizada para realizar análisis de sentimiento. En el apéndice se exhiben distintas formas funcionales de cálculo de la ponderación global.

Ahora bien, con respecto al factor de normalización se basa en dos aspectos:

1. Los documentos más largos contienen términos más distintos: esto influye significativamente en la similitud entre ellos porque existen más coincidencias en las dimensiones individuales en el espacio vectorial.
2. La frecuencia de términos en documentos más largos es mayor.

Para su cálculo, los pesos del vector se combinan en un solo número. Estos pesos son los pesos locales multiplicados por los pesos globales. A los fines del presente trabajo se utiliza la técnica de normalización del coseno por ser la más utilizada y tener la ventaja de proveer mayor robustez estadística a los datos. Esta normalización transforma simplemente un vector en un vector unitario (un vector de longitud 1) dividiéndolo por su longitud. Esto aborda simultáneamente ambas razones para la normalización de la longitud del documento: altas frecuencias y un mayor número de términos (elementos distintos de cero) en un documento. Por lo tanto, el factor de normalización tiene un valor más alto. En el apéndice 2 se exhiben distintas formas funcionales del factor de normalización.

Luego de la etapa de preprocesamiento de los datos y del ordenamiento en vectores con la técnica de *Bags of Words*, se realiza la medición del sentimiento de cada uno de los tweets recolectados en la ventana analizada. El análisis de sentimiento se medirá mediante el cálculo de la polaridad de cada documento procesado en el análisis de texto realizado.

En el presente trabajo se realiza una medición de la polaridad para calcular el indicador de sentimiento que posteriormente se emplea como regresor en el modelo lineal. La metodología empleada para la polarización fue propuesta por Feuerriegel *et al.* (2019), y fue seleccionada respecto a otras implementaciones por las siguientes particularidades: i) cuenta con un diccionario financiero, sumamente importante para el tratamiento de las palabras vacías (StopWords), ii) modela la variable respuesta (la polaridad de cada documento, en este caso, cada *tweet*) como una variable continua, es decir rompe con la lógica dicotómica y permite que la polaridad asuma un valor continuo y iii) la implementación en la librería utilizada en R se puede adaptar fácilmente a distintos idiomas, como por ejemplo el español.

Estos autores consideran a cada *tweet* como una observación, es decir emplea a cada una de las palabras de un *tweet* (o documento) como variables explicativas del mismo²¹. Es decir,

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j x_{ji} + \epsilon_i \quad (i = 1, 2, \dots, n) \quad (9)$$

Donde,

- y_i representa la polaridad del i -ésimo *tweet*.
- x_{ji} ($j = 1, 2, \dots, p$) representa la cantidad de veces que aparece la palabra individual j en el i -ésimo *tweet* procesado.
- β_0 es el intercepto de la regresión lineal múltiple del i -ésimo *tweet*.

²¹Esta metodología se encuentra implementada en R en la librería *SentimentAnalysis*. Dentro de este paquete se utilizó la función *analyzeSentiment* para calcular la puntuación de polaridad de cada palabra procesada en el tweet. Este paquete proporciona una colección de herramientas para realizar y comparar diferentes métodos para el análisis de sentimientos. Además, esta implementación cuenta con varias ventajas sobre otras librerías del mismo lenguaje, dentro de las cuales podemos encontrar un diccionario financiero desarrollado por Loughran y McDonald's (Loughran *et al.*, 2011), que se puede personalizar mediante la selección de varios enfoques. Asimismo, la variable respuesta del modelo es de naturaleza continua. Además, otra ventaja de esta librería radica en que se puede adaptar fácilmente a distintos idiomas, como por ejemplo el español.

- β_j ($j = 1, 2, \dots, p$) mide el efecto de la palabra j en la variable respuesta del i -ésimo *tweet*.
- ϵ_i indica la perturbación aleatoria del modelo propuesto en el i -ésimo *tweet*.

Los coeficientes β_j reflejan la polaridad de cada una de las palabras preprocesadas en cada *tweet*. Es decir, las magnitudes de los coeficientes de regresión miden la percepción estadística de las palabras individuales de cada *tweet*.

La metodología de Feuerriegel *et al.* (2019) utilizada en este trabajo, incorpora en la implementación un término de penalización en la función de pérdida empleada para estimar los coeficientes del modelo de regresión lineal múltiple. Este término se utiliza para penalizar (o castigar) la incorporación de nuevos regresores con la finalidad de optimizar la selección de aquellas palabras relevantes en cada *tweet* procesado.

La penalización empleada en esta metodología se conoce como regularización LASSO y es una herramienta útil para extraer aquellas palabras decisivas del contenido de cada *tweet* procesado²². A los fines de este trabajo, esta implementación ha permitido descartar aquellas palabras que para un *tweet* en particular tenían una baja ocurrencia de aparición con respecto al resto de las otras palabras analizadas, dado que en la minimización de cuadrados propuesta fuerza a que los coeficientes tiendan a cero. La función de pérdida utilizada en esta implementación se formaliza en la siguiente expresión.

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (10)$$

El primer término de la expresión anterior refleja la suma de los residuos al cuadrado y el segundo a la penalización, que representa la suma del valor absoluto de los coeficientes de regresión. Asimismo, el grado de penalización está controlado por el hiperparámetro λ y cuando es igual a cero, entonces el resultado equivale a un modelo lineal estimado por mínimos cuadrados ordinarios (MCO). A medida que λ aumenta, mayor es la penalización y más predictores quedan excluidos. En el presente trabajo este hiperparámetro se obtiene mediante la utilización de una grilla con un rango acotado de posibles valores a asumir²³.

Como en cada día de recopilación de tweets, se dispone de una cantidad distinta de documentos (m_t) ($t = 1, 2, \dots, n$), se procedió a resumir en un único valor el sentimiento con la finalidad de generar una serie diaria de información, que pueda ser utilizada como un regresor en una especificación que se planteará en el apartado Metodología propuesta del Capítulo 3.

Para cada *tweet* se procedió a estimar mediante la especificación de la ecuación (9) un valor de polaridad y para todos los *tweets* correspondientes a un mismo día de análisis t se resumió la información en el valor mediano de los valores estimados para un mismo día para obtener un valor diario del indicador del sentimiento diario. La utilización del valor mediano de las estimaciones realizadas con la ecuación (9) para cada uno de los días de la ventana analizada, se emplea porque dicha medida de tendencia central no se encuentra

²²Esta regularización permite que una variable explicativa que posee un coeficiente de regresión igual a cero no influya en el modelo, y de esta forma se excluyen aquellos predictores menos relevantes, priorizando que los menos influyentes queden excluidos.

²³Se utiliza un enfoque Grid Search para obtener el valor óptimo del hiperparámetro λ .

afectada por la presencia de valores atípicos o extremos, a diferencia de otras medidas, como por ejemplo la media aritmética.

En la siguiente ecuación se exhibe el planteo formal de obtención del valor diario del sentimiento,

$$Sentiment_t = \hat{y}_t \left(\frac{m_t + 1}{2} \right) \quad (11)$$

Donde la expresión anterior indica que el valor del sentimiento $Sentiment_t$ para el día de observación t se obtiene como la mediana reflejada como aquel valor de los valores estimados para un mismo día t ubicada en el orden $\left(\frac{m_t+1}{2}\right)$, donde m_t indica la cantidad de *tweets* recopilados para el día t .

De esta forma, se resuelve en este trabajo la medición diaria del sentimiento sobre las opiniones de los comentarios de *Twitter*. Asimismo, es importante destacar que los valores estimados con esta implementación arrojan valores de la variable respuesta de naturaleza continua, lo que representa una novedad de este trabajo.

2.3. Los modelos de NLP en el análisis de sentimiento aplicado en contextos económicos y financieros

El análisis de sentimiento y de expectativas en contextos de redes sociales, como *Twitter* se evidencia en distintos trabajos académicos que ponen de relieve la información valiosa que es posible extraer de distintos medios sociales.

Por ejemplo, en el trabajo de Aromí *et al.* (2022), utilizan el contenido de *Twitter* para generar un indicador del nivel de atención asignado a la inflación en las discusiones públicas, y lo comparan en forma favorable con otros indicadores como el contenido de los medios, los tweets de los medios, la intensidad de búsqueda en Google y las encuestas de consumidores. Es importante destacar que los modelos de pronóstico estimados muestran que el indicador proporciona información valiosa sobre los niveles futuros de inflación. Ambos autores concluyen como una futura mejora lo siguiente,

“Finally, this study evaluated regularities using monthly time series. Analyses at higher frequencies can provide further insights regarding the relationship between social media content and inflation dynamics” (Aromí *et al.*, 2022, pp. 17)

Es decir, consideran que un análisis de frecuencia más alta, como puede ser la diaria como la llevada a cabo en este trabajo, puede brindar información valiosa en los datos respecto a la relación entre el contenido de las redes sociales y la dinámica de la inflación.

Asimismo, Aromí (2022) en un trabajo anterior propone un índice que describe las opiniones económicas transmitidas por usuarios argentinos en la red social *Twitter* y concluye que la utilización de técnicas de procesamiento de lenguaje natural (NLP²⁴) constituye un elemento clave para el desempeño satisfactorio del indicador de opiniones que ha construido. Una de los aspectos claves de este trabajo resulta en la utilización

²⁴Acrónimo en inglés de *Natural Language Processing*

de lista de palabras clave para identificar mensajes relacionados con temas económicos, la cual se genera mediante un modelo de detección de temas entrenado con mensajes compartidos en *Twitter* por usuarios argentinos. Esto deja en claro la importancia de realizar un adecuado procesamiento de los datos. En el presente trabajo se emplea el modelo de Feuerriegel *et al.* (2018) porque contiene embedido un diccionario de palabras financieras que es sumamente útil para el contexto del problema planteado.

Asimismo, Arias *et al.* (2012) se proponen evaluar de manera rigurosa, si un indicador de sentimiento público extraído de los mensajes diarios de *Twitter* puede mejorar la previsión de indicadores sociales, económicos o comerciales. Lo primordial de este trabajo se ubica en la conjetura formulada respecto a que, si *Twitter* ayuda, entonces esto debería reflejarse en el hecho de que las predicciones de los modelos que usan datos relacionados con *Twitter* son mejores que los modelos que no usan estos datos. Esto es lo que los autores han querido dejar en evidencia a través de su artículo, que al incorporar información procesada de *Twitter* mediante indicadores de sentimientos los modelos deberían mejorar sustancialmente sus predicciones.

En esta línea, Kumar *et al.* (2023) emplean indicadores de sentimientos para pronosticar la evolución del mercado de valores Dow Jones basándose en las ideas del trabajo de Arias *et al.* (2012), donde evidencian que la extracción de artículos de noticias y la predicción de los movimientos de los precios de los productos en función del contenido del corpus de reseñas se vuelve beneficiosa.

Con respecto a la correlación entre las predicciones en el mercado de acciones y el sentimiento de las opiniones extraídas de *Twitter*, existen varios trabajos (Deveikyte *et al.*, 2020; Kordonis *et al.*, 2017; Kukk, 2018) con enfoques similares.

Por ejemplo, Kukk (2018) investigó si categorizar los *tweets* en un mayor número de categorías (ira, disgusto, alegría, sorpresa, ninguna) resulta en correlaciones más fuertes y concluyó que el sentimiento de los *tweets* mostró correlaciones más fuertes con la rentabilidad de las acciones que la emoción, aunque el tipo de correlación encontrada difirió entre las empresas consideradas. Por tal motivo, sugieren que la división de los *tweets* en menos categorías daría como resultado etiquetas semánticamente más distintas que son más fáciles de distinguir y que, por lo tanto, muestren correlaciones más fuertes.

Asimismo, Deveikyte *et al.* (2020) exhiben un análisis donde abordan la predicción de la volatilidad del mercado mediante una aproximación con un análisis de sentimiento. Según dichos autores los hallazgos sugieren que existe evidencia de correlación entre el sentimiento y los movimientos del mercado de capitales, es decir el sentimiento capturado en los titulares de las noticias podría emplearse como una señal para predecir los rendimientos del mercado.

Además, Kordonis *et al.* (2016) investigaron si el sentimiento público, medido a partir de *tweets*, se correlaciona o si puede emplearse como predictor del precio de las acciones, específicamente para 16 de las empresas de tecnología más populares. Una de las conclusiones arribadas es que los cambios en la opinión pública pueden afectar al mercado de capitales, lo que interpretan que puede utilizarse para predecir el mercado de valores con elevadas probabilidades.

Perales González *et al.* (2017) afirman que las series de tiempo vinculadas a opiniones sobre una institución financiera podrían correlacionarse con las series financieras.

“Searching for tweets of a banking institution and comparing not the absolute

values but the variations between days or weeks lead us to evaluate how the reputation of this institution evolves. This variation can be associated to direct actions this institution does, so an strategy to avoid bad reputation and look for good reputation can be advised. Those time series can be correlated with financial series. This information could be useful in order to measure economic impact due to reputation. Keeping in mind, the latest goal of reputational analysis is to quantify reputational risk and this cannot be fulfilled without a measure of the money that is in risk(...)" (Perales González et al., 2017, p.11)

En el presente trabajo se estudiará la correlación entre el sentimiento negativo de comentarios de *Twitter* sobre una Entidad Financiera y el rendimiento logarítmico diario de la acción en el mercado de valores donde cotiza. La finalidad es determinar el Capital Económico por Riesgo Reputacional.

Resumen de Capítulo

En el presente capítulo se profundizó en la utilización de técnicas de *Text Mining*, en particular de *Sentiment Analysis* para medir el Riesgo Reputacional de Empresas en general y en particular de Entidades Financieras.

Por tal motivo, para cumplimentar con dicho objetivo se describieron y analizaron varios puntos, entre ellos el concepto de *Big Data*, Análisis de Sentimiento, el modelo de *Bag-of-Words* y la metodología de medición del sentimiento que se utiliza en este trabajo.

Por último, se realizó la articulación entre el Análisis de Sentimiento y la medición del Riesgo Reputacional mediante la exhibición de varios artículos académicos que concluyen sobre la presencia de correlaciones significativas entre el sentimiento de las opiniones públicas y el comportamiento del precio de los activos. El siguiente capítulo exhibe una aplicación concreta de estas técnicas para determinar el Capital Económico del Banco Santander Argentina.

"...Lo importante no es llegar, lo importante es el camino; yo no busco la verdad, sólo sé que hay un destino..."

Fito Páez, *Eso que llevas ahí*

Capítulo 3: Aplicación de Text Mining para medir el Riesgo Reputacional

En este capítulo se exhiben los resultados de la aplicación del Análisis de Sentimiento en la medición del Riesgo Reputacional de Entidades Financieras para hacer frente a eventos reputacionales que afecten la estabilidad de la empresa.

Para cumplimentar con dicho objetivo, se utilizarán datos extraídos de *Twitter* vinculados a opiniones de usuarios de la red social sobre el Banco Santander Argentina S.A.. Primero, se realizará un análisis descriptivo de los datos. Luego se exhibirán los resultados del cálculo del indicador de sentimiento que se empleará para medir el riesgo reputacional calculado con el modelo de polarización descrito en el capítulo previo y se realizará una estimación del Capital Económico que tendrá que reservar dicha Entidad Financiera. Por último, se esbozarán algunas recomendaciones propuestas para la gestión de este riesgo.

3.1. Análisis Exploratorio de Datos

En el presente trabajo se utilizó la API gratuita de la red social *Twitter* para acceder a datos de opiniones públicas realizadas sobre el Banco Santander Argentina S.A. que cotiza en el Mercado de Valores de Buenos Aires, también conocido como BYMA (Bolsas y Mercados Argentinos)²⁵.

En el mercado local cotizan las siguientes empresas pertenecientes a la industria bancaria o financiera²⁶.

²⁵Se utilizó la función `searchTwitter` de la librería `twitteR` para extraer tweets. Documentación en <https://www.rdocumentation.org/packages/twitteR/versions/1.1.9/topics/searchTwitter>. Esta aplicación permite solamente extraer *tweets* de los últimos 10 días. En las conclusiones se esboza como se resolvió esta limitación presupuestaria.

²⁶Se extrajo de la página oficial del BYMA <https://www.byma.com.ar/emisoras/empresas-listadas/>.

Razón Social	Especie (Ticker)
Banco Hipotecario S.A.	BHIP
Banco Macro S.A.	BMA
Banco Patagonia S.A.	BPAT
Banco Santander Argentina S.A.	BRIO
BBVA Banco Francés S.A.	BBAR
Grupo Financiero Galicia S.A.	GGAL
Banco de Valores S.A.	VALO
Grupo Supervielle S.A.	SUPV

Tabla 1. Denominación en BYMA

La tabla anterior exhibe las Entidades Financieras privadas más importantes de nuestro país según el ranking que desarrolla el Banco Central de la República Argentina (BCRA). Este ranking se realiza en función de los niveles de Depósitos, de Activos, de Pasivos y de Patrimonio Neto (PN)²⁷. En la siguiente tabla se exhibe el ranking a Marzo 2021 (último dato disponible a Febrero 2022) desarrollado por el BCRA en función del nivel de Activos.

Razón Social	Activos	Posición (Activos)
Banco Hipotecario S.A.	\$ 132.789.989	18
Banco Macro S.A.	\$ 746.144.409	5
Banco Patagonia S.A.	\$ 320.366.325	12
Banco Santander Argentina S.A.	\$ 1.078.178.109	3
BBVA Banco Francés S.A.	\$ 727.821.641	6
Grupo Financiero Galicia S.A.	\$ 1.048.438.288	4
Banco de Valores S.A.	\$ 110.596.683	19
Grupo Supervielle S.A.	\$ 277.950.353	14

Tabla 2. Ranking de Activos - Sistema Financiero

Es importante destacar que los montos están expresados en miles de Pesos Argentinos. La posición relativa a un total de 80 (ochenta) entidades financieras y las primeras ubicaciones pertenecen a empresas de titularidad estatal (gubernamental), tales como el Banco de la Nación Argentina (BNA) y el Banco de la Provincia de Buenos Aires²⁸. En el presente trabajo se ha decidido realizar la aplicación práctica con el Banco Santander Argentina por ser la entidad financiera de capitales privados más importante del país y representar una de las empresas del sector consideradas de riesgo sistémico por la autoridad monetaria. En los siguientes gráficos se exhibe la evolución del precio de cierre ajustado (*Adjusted Price*) de cada una de las Entidades Financieras cotizantes desde Enero de 2018 a Mayo 2022²⁹.

²⁷Entre dichas instituciones figuran algunas consideradas de Riesgo Sistémico (D-SIB), como por ejemplo el Banco Galicia y el Banco Macro, las cuales deben cumplir con ciertos requerimientos de información por parte de la autoridad monetaria en función de lo solicitado por el Comité de Supervisión Bancaria de Basilea.

²⁸El banco forma parte del grupo económico BAPRO.

²⁹El dato más actualizado se corresponde al Viernes 6 de Mayo de 2022.

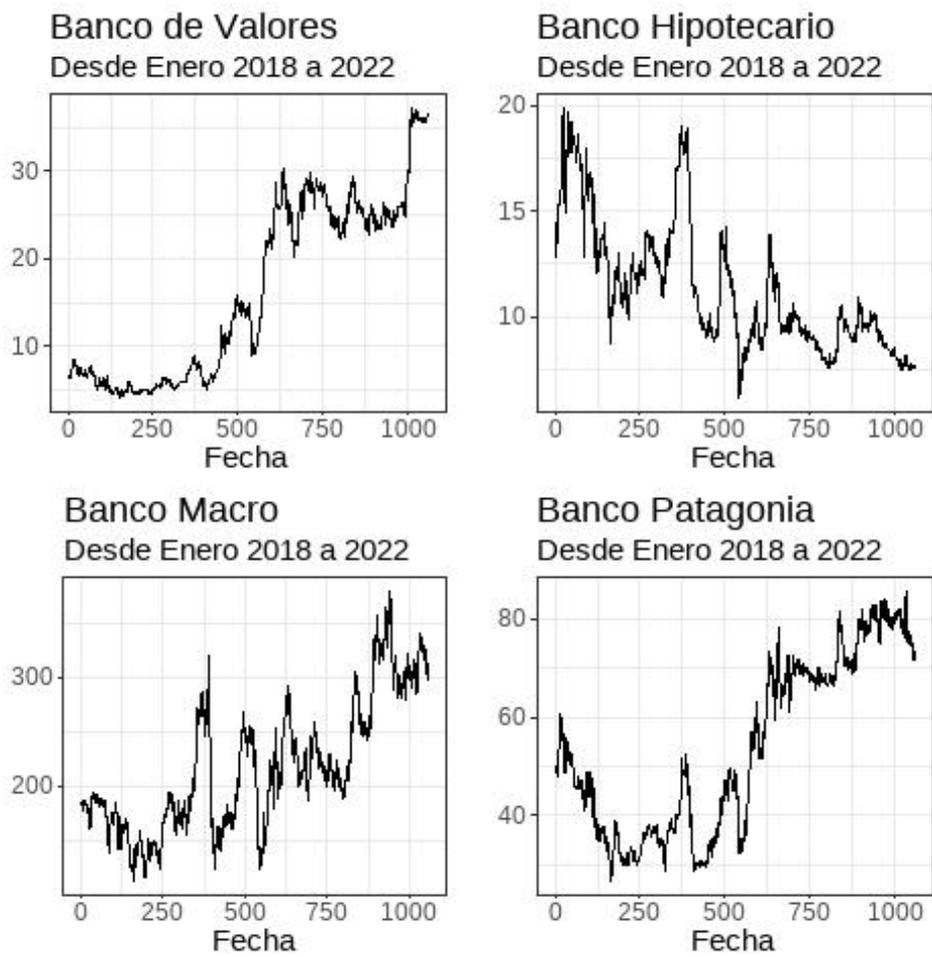


Figura 7. Evolución de la cotización de cada Acción

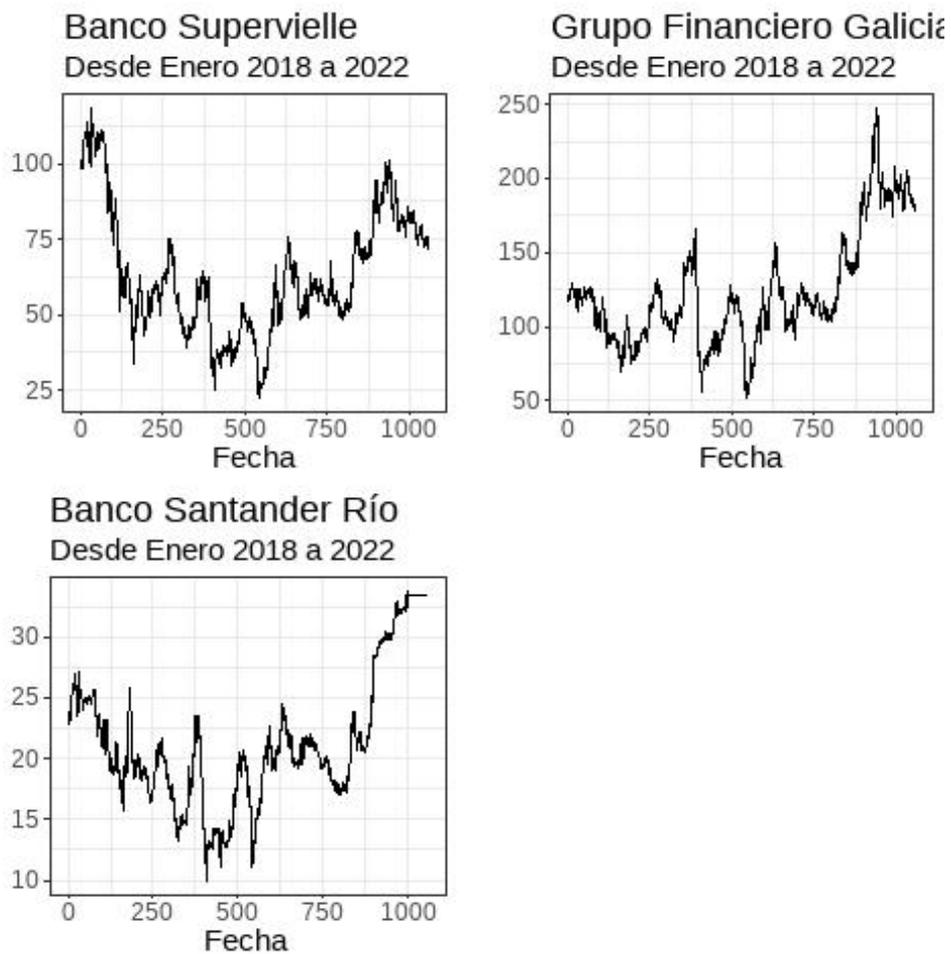


Figura 8. Evolución de la cotización de cada Acción

En el siguiente gráfico se observa la evolución de los retornos nominales acumulados (en términos porcentuales) desde Enero de 2018 a Mayo 2022. Esto permite entender como se hubiese comportado una inversión realizada el primer día hábil de Enero de 2018 manteniendo la misma hasta la actualidad³⁰. Se observa que una inversión en Banco de Valores, Banco Macro, Banco Patagonia, Grupo Financiero Galicia y en menor medida Banco Santander Argentina han obtenido rendimientos acumulados positivos, y en algunos casos han duplicado la inversión inicial. Sin embargo, este rendimiento fue inferior a la inflación acumulada de Argentina en el período analizado. Asimismo, se evidencia que los retornos acumulados de cada una de las acciones disminuyeron fuertemente para los primeros meses del 2020, que se corresponde con el inicio de la pandemia de COVID-19³¹.

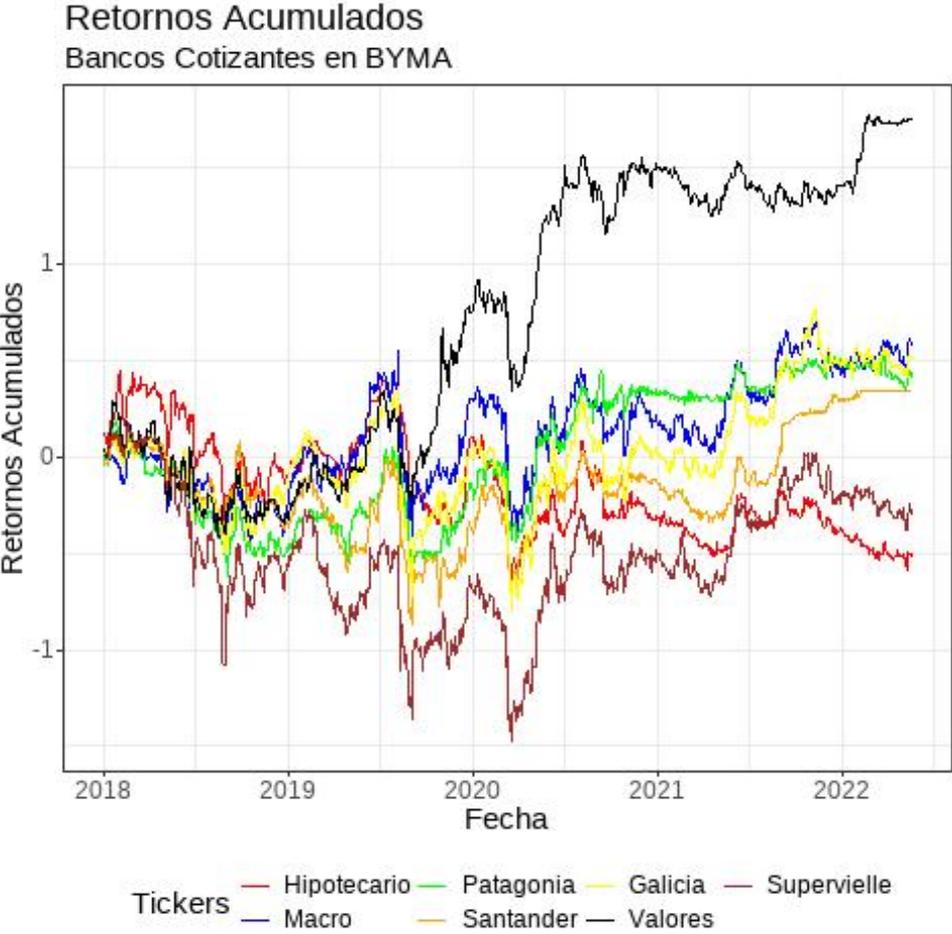


Figura 9. Evolución de los Retornos Acumulados

En el siguiente gráfico se exhibe un histograma sobre los retornos diarios de las acciones de cada Entidad Financiera.

³⁰Lo que en la jerga financiera se denomina *buy and hold*.

³¹Acrónimo en inglés de Coronavirus Disease 2019.

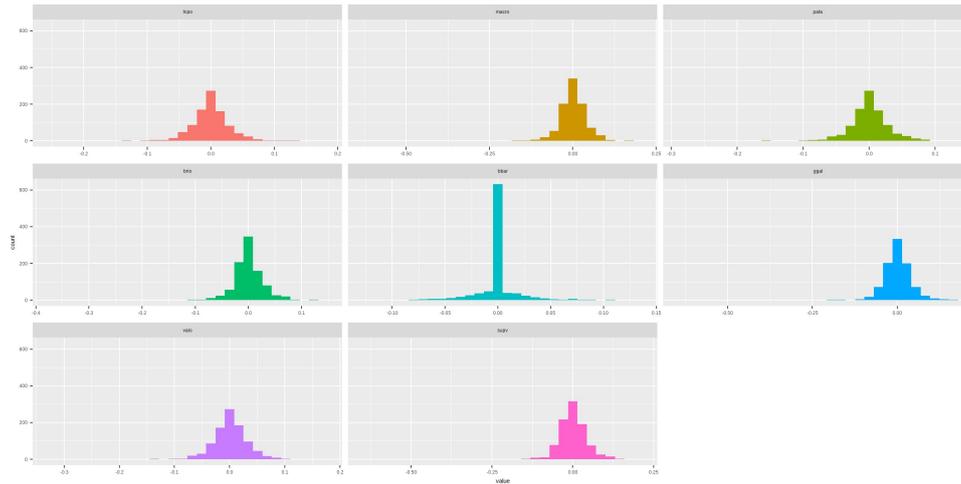


Figura 10. Histogramas de los Retornos Diarios

Estas series financieras en el período analizado se caracterizan por su forma leptocúrtica y la presencia de colas pesadas (o gruesas). En la siguiente tabla se resumen ciertas medidas descriptivas sobre el conjunto de datos.

Medida	BHIP	BMA	BPAT	BRIO	BBAR	GGAL	VALO	SUPV
Media	-0,000336	0,000574	0,000495	0,000242	-0,000058	0,000582	0,00143	-0,000153
Desvío	0,032976	0,044695	0,030682	0,030683	0,021673	0,043794	0,032765	0,048487
Mediana	0	0,000293	0	0	0	-0,000744	0	-0,001617
Mínimo	-0,247177	-0,62578	-0,28636	-0,378436	-0,121039	-0,636084	-0,324518	-0,650228
Máximo	0,180436	0,188031	0,124808	0,13276	0,136132	0,164128	0,16791	0,189931
Varianza	0,001087	0,001998	0,000941	0,000941	0,00047	0,001918	0,001074	0,002351
Asimetría	-0,582502	-2,954953	-0,994407	-1,865704	0,418661	-3,330382	-1,017081	-2,909209
Curtosis	7,587353	41,50331	10,145902	26,395937	9,134384	48,045488	11,93068	37,664286
n	1001	1001	1001	1001	1001	1001	1001	1001

Tabla 3. Resumen de medidas descriptivas

El retorno promedio diario de cada uno de los activos financieros analizados es cercano a cero, representa una característica habitual de este tipo de series financieras. El desvío estándar de los retornos diarios (o volatilidad diaria) representa una primera medida del riesgo precio (o de mercado) del activo financiero. En todos los casos, la curtosis es mayor a 3 lo que indica que los retornos diarios exhiben un exceso de curtosis respecto a la normal (leptocurtosis). En el apéndice de Gráficos se encuentran otras visualizaciones sobre los retornos de las acciones de los bancos cotizantes en el BYMA en el período bajo análisis, los cuales complementan al análisis realizado en este apartado.

3.2. Aplicación de Sentiment Analysis al cálculo del Capital Económico

En este apartado se exhiben los resultados de la aplicación de la técnica de *Text Mining* denominada Análisis de Sentimiento para realizar una aproximación al cálculo del Capital Económico por Riesgo Reputacional para el Banco Santander Argentina. Para cumplimentar con lo mencionado, en primera instancia se esboza la propuesta metodológica utilizada para la aplicación práctica. Luego se expone información sobre el indicador de reputación calculado con el modelo de polaridad descrito en el capítulo anterior. Por último, se exhiben los resultados del modelo estimado y se explica el procedimiento para realizar la prueba retrospectiva (*Backtesting*)³².

3.2.1. Metodología propuesta

En este trabajo se utilizará un enfoque similar al exhibido en el apartado 1.3.1 (En función de la volatilidad del precio de las acciones) mediante el empleo de un modelo APT de riesgo reputacional. En este modelo se emplea un indicador de reputación de naturaleza continua calculado con el sentimiento que surge de opiniones públicas de usuarios de *Twitter*. La finalidad es relacionar el retorno (rendimiento) R_t del activo de la Entidad Financiera al momento t con el sentimiento que surge de la aplicación de la técnica de *Text Mining*, $Sentiment_t$. Además se estiman otras tres especificaciones para incorporar variables de control. A continuación se exhiben las cuatro especificaciones del modelo.

$$R_t = \alpha + \beta_1 Mkt \text{ Return}_t + \epsilon_t \quad (12)$$

$$R_t = \alpha + \beta_1 Sentiment_t + \epsilon_t \quad (13)$$

$$R_t = \alpha + \beta_1 Sentiment_t + \beta_2 Mkt \text{ Return}_t + \epsilon_t \quad (14)$$

$$R_t = \alpha + \beta_1 Sentiment_t + \beta_2 Mkt \text{ Return}_t + \beta_3 Fcial \text{ Mkt Return}_t + \epsilon_t \quad (15)$$

La primera especificación (ecuación 12) relaciona el retorno diario de la acción del Banco Santander Argentina con el retorno del índice bursátil *S&P Merval*. Donde el coeficiente de regresión β_1 mide la sensibilidad de los retornos diarios de la empresa a las variaciones de los retornos diarios del mercado.

La segunda especificación (ecuación 13) relaciona el retorno diario de la acción del Banco Santander Argentina con el indicador de sentimiento diario calculado por análisis de sentimiento. En esta ecuación el coeficiente de interés β_1 refleja la sensibilidad de R_t a las variaciones de $Sentiment_t$.

En la tercera especificación (ecuación 14) se incorpora a la ecuación (13) el retorno diario del índice bursátil *S&P Merval* como variable de control para el análisis realizado y por último, en la cuarta especificación (ecuación 15) se incorpora a la ecuación (14) el retorno diario del índice bursátil sectorial financiero *S&P Merval Financiero* como otra variable

³²La prueba retrospectiva no se realizó porque la empresa dejó de cotizar en dos mercados, entre ellos BYMA. Esta decisión fue tomada por los accionistas que aprobaron retirar las acciones de los mercados de capitales de Argentina y España. Por lo tanto, al no disponer del precio de la acción no es posible validar el modelo en una ventana de tiempo más alejada.

de control³³. Asimismo, los otros términos de cada una de las ecuaciones formuladas representan,

- α es el intercepto
- ϵ_t es el término de error de la acción en el tiempo t

El Capital Económico (en Riesgo) por riesgo reputacional ($Capital_{Rep}$) se calcula mediante la siguiente ecuación,

$$Capital_{Rep} = V_t * \beta_{Sentiment} * P_{95\%} \quad (16)$$

El primer factor de la ecuación anterior representa el valor de mercado al momento t del Banco Santander Argentina (V_t) resultante del producto entre el precio de cierre ajustado de la acción de P_t y la cantidad de acciones en circulación de valores Q_t . Formalmente,

$$V_t = P_t * Q_t \quad (4)$$

El segundo factor $\beta_{Sentiment}$ representa el coeficiente de regresión, que estima el efecto del sentimiento expresado en las opiniones de los usuarios. Por último, el tercer factor $P_{95\%}$ representa el percentil del indicador de la reputación para considerar valores extremos con una probabilidad de 95 % de ocurrencia. Es importante aclarar que el nivel de confianza se eligió en base a prácticas habituales de mercado para medir la peor pérdida en condiciones normales de mercado. Ahora bien, este parámetro de la metodología podría ser calibrado según el nivel de tolerancia y el apetito al riesgo de la entidad bajo análisis.

El monto determinado mediante la ecuación (16) representa el capital que una Entidad Financiera debería reservar para hacer frente a eventos reputacionales poco probables pero posibles de ocurrencia, que pueden afectar la solvencia de la firma. Es importante dejar en claro que dicho guarismo no representa el capital regulatorio³⁴.

3.2.2. Text Mining sobre los datos

Para cumplimentar con el objetivo propuesto se trabajó con los datos del Banco privado más importante de la República Argentina, en función del Ranking de Entidades Financieras publicado por el Banco Central de la República Argentina.

Por lo tanto, se recolectaron *tweets* a partir de ciertas palabras denominadas “claves”, con las cuales los usuarios de la red social pudiesen realizar comentarios sobre la entidad financiera en cuestión. Las *keywords* empleadas han sido,

1. #bancosantanderrio
2. #bancosantander
3. bancosantander
4. bancosantanderrio
5. @bancosantander

³³<https://www.spglobal.com/spdji/es/documents/methodologies/methodology-sp-byima-indices-spanish.pdf>.

³⁴En la sección 1.4.1. se hace referencia a este capital.

6. @Santander_Ar

Es importante dejar en claro que se han elegido dichas *keywords*, luego de un proceso de verificación de aquellas palabras y combinaciones más utilizadas por los usuarios de la red social para hacer mención sobre la organización. Ahora bien, se han probado distintas palabras, como así combinaciones de las mismas para determinar cuáles son más adecuadas para realizar el análisis. Se entiende que es una de las cuestiones principales al momento de armar la base de datos, tener en claro aquellas palabras o cuentas que los usuarios de la red social emplean para mencionar a la firma bajo análisis³⁵.

Se comenzó a recolectar datos desde el 6 de Enero de 2022 y se finalizó el 7 de Febrero de 2022 para contar con un mes de extracción diaria de datos. La extracción del 6 de Enero dispone de datos de la red social desde el 29 de Diciembre de 2021 hasta dicho día inclusive. Por lo tanto, en cada día de recolección se disponía de datos de días anteriores, que al unificarlos se presentaron datos repetidos que fueron descartados. De esta forma se aseguró contar con la totalidad de datos provenientes del mes de Enero de 2022 sin duplicar información. Por lo tanto, la base total de datos cuenta con aproximadamente 165.000 *tweets* únicos (registros/filas) con 87 variables en 32 días de recolección.

Luego se procedió a realizar un análisis descriptivo y visual del conjunto de datos. Por ejemplo, en la siguiente imagen se exhibe la cantidad de menciones por hora sobre el Banco Santander Argentina.

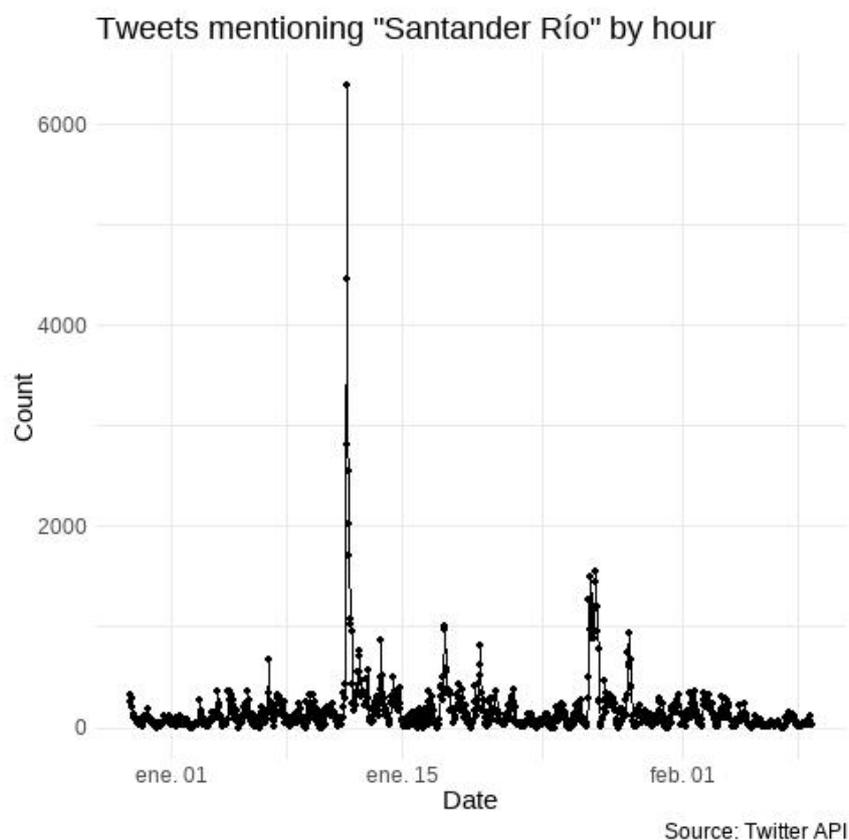


Figura 11. Cantidad de Menciones por hora

³⁵Los *hashtags* son un conjunto de caracteres precedidos por un el símbolo numeral (#) que sirve para identificar o etiquetar un mensaje en las webs de microblogs.

En el gráfico anterior se observa la presencia de ciertos días con menciones por encima al valor promedio de la serie utilizada. No obstante, de estos datos se aplicaron ciertos procedimientos explicados en el capítulo anterior para utilizarlos en el algoritmo de Análisis de Sentimiento³⁶. Asimismo, es posible graficar la densidad de la cantidad de *Tweets* por fecha y por hora.

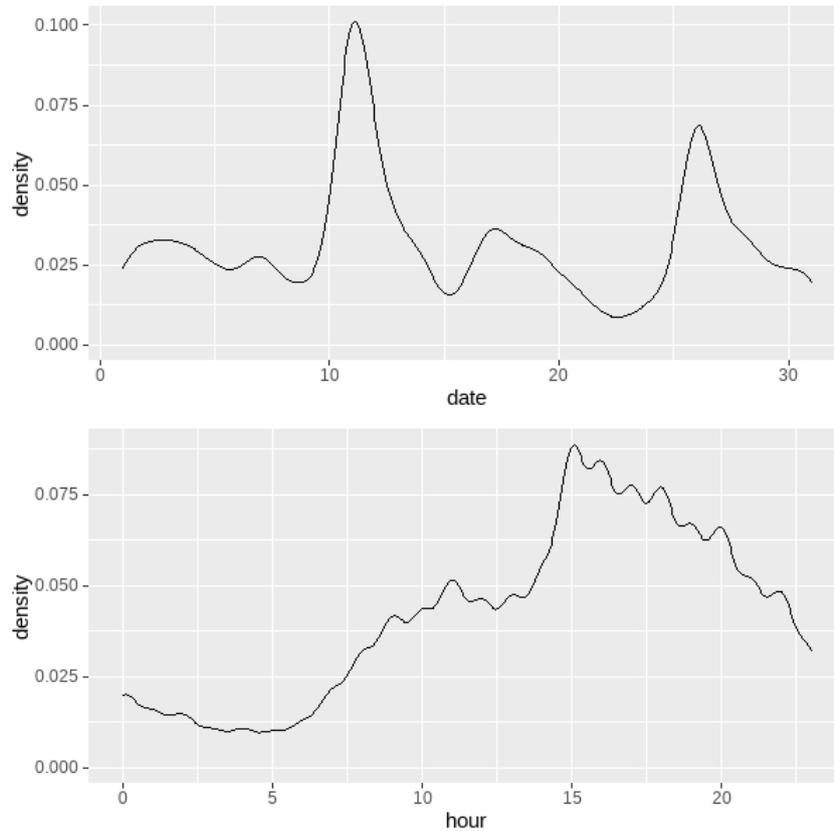


Figura 12. Densidad por fecha y por hora

Del gráfico anterior se observa que ciertos días del mes de Enero 2022 poseen más densidad, como por ejemplo el 12 de Enero y esto se debe a que se han realizado más menciones a los términos buscados. Esta visualización permite entender si existió algún evento coyuntural externo que haya generado una mayor mención en la red social, como por ejemplo algún servicio caído como el *Home Banking*, o alguna noticia vinculada al lavado de activos y financiamiento del terrorismo.

Luego de haber realizado el proceso de transformación de los datos, mediante la limpieza y la sanidad de la base se procedió a calcular la medida diaria del sentimiento de los usuarios de *Twitter* hacia la organización a partir de las palabras, *hashtags* y cuentas buscadas. Este indicador fue calculado mediante la metodología de Feuerriegel *et al.* (2019) descrito en el apartado 2.2. de este trabajo. En el siguiente gráfico se exhibe la evolución diaria del indicador de sentimiento sobre *tweets* vinculados a la organización Santander Argentina S.A.,

³⁶Los procedimientos empleados son filtrado de *stop words*, términos comunes y raros, normalización, tokenización y detección de oraciones, entre otros.

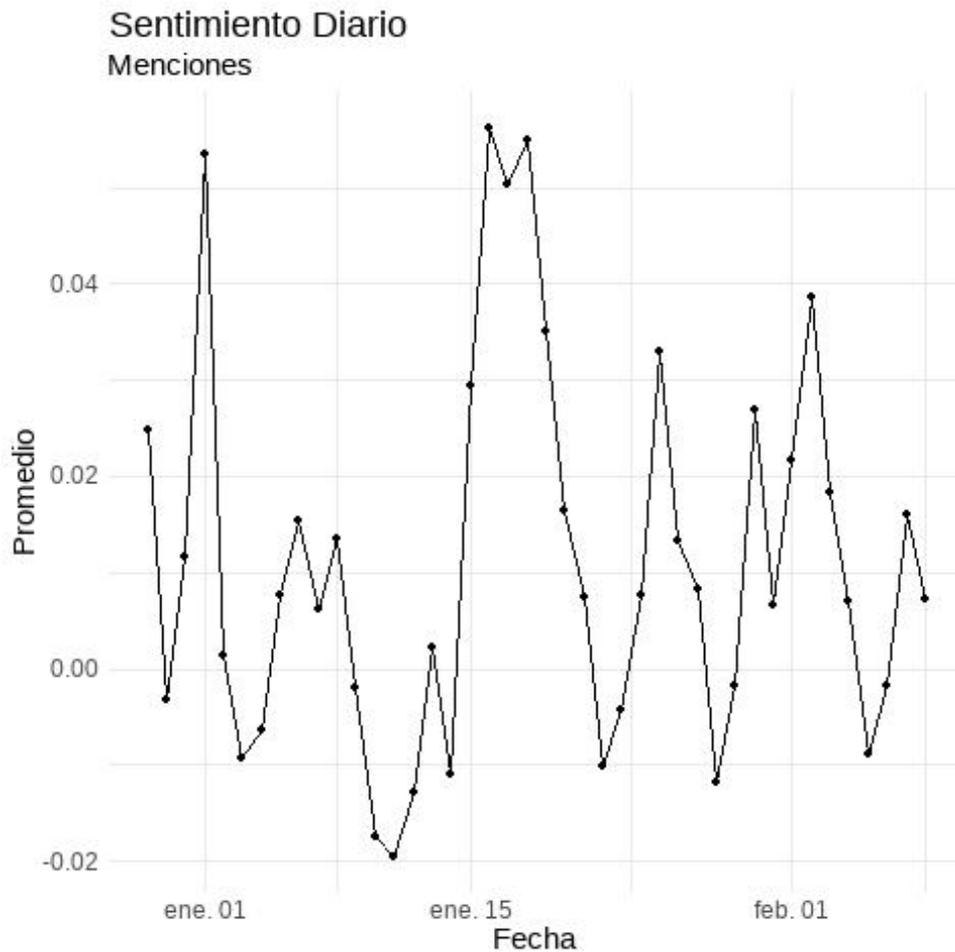


Figura 13. Sentimiento Diario de Menciones

Este indicador diario del sentimiento sobre las menciones de usuarios de la red social sobre el Banco Santander Argentina incluye todas las menciones, ya sean de índole negativa o no. Por lo tanto, en el gráfico anterior se observa que el indicador determinado asume en algunos días un valor positivo, lo que refleja que el sentimiento de los *tweets* de ese día han proporcionado una polaridad positiva sobre la compañía. En el apéndice se exhiben otros gráficos de evolución de sentimientos diarios relacionados a este indicador, como por ejemplo la evolución del valor promedio diario, el total de menciones y distintas medidas de variabilidad. En el siguiente gráfico se exhibe el histograma del indicador diario del sentimiento y se observa la forma con asimetría negativa de los datos, que coincide con el coeficiente estandarizado de asimetría exhibido en la tabla 4. Asimismo, el valor medio de este indicador es 0,015828445 con un desvío de 0,018860726 y el valor mediano es 0,017261905.

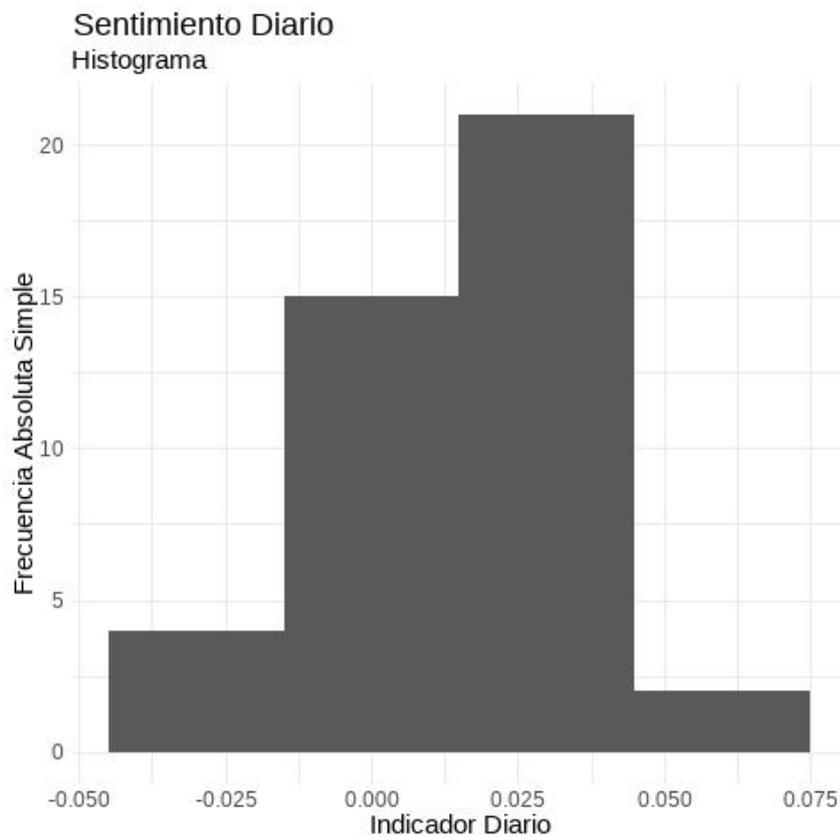


Figura 14. Histograma del Sentimiento Diario

En la siguiente tabla se resumen ciertas medidas descriptivas sobre el indicador diario del sentimiento.

Medida	Sentiment
Media	0.015828445
Desvío	0.018860726
Mediana	0.017261905
Mínimo	-0.029411765
Máximo	0.071428571
Varianza	0.000355727
Asimetría	-0.217914590
Curtosis	1.392268563
Percentil 5	-0.02263258
Percentil 95	0.04133826
n	42

Tabla 4. Resumen de medidas descriptivas

3.2.3. Resultados del modelo estimado

La relación entre el sentimiento de los *tweets* y el rendimiento logarítmico diario del Banco Santander Argentina exhibe una relación inversa entre ambas variables. En el siguiente gráfico se exhibe la nube de puntos entre ambas variables y la recta de regresión estimada en base a la especificación seleccionada.

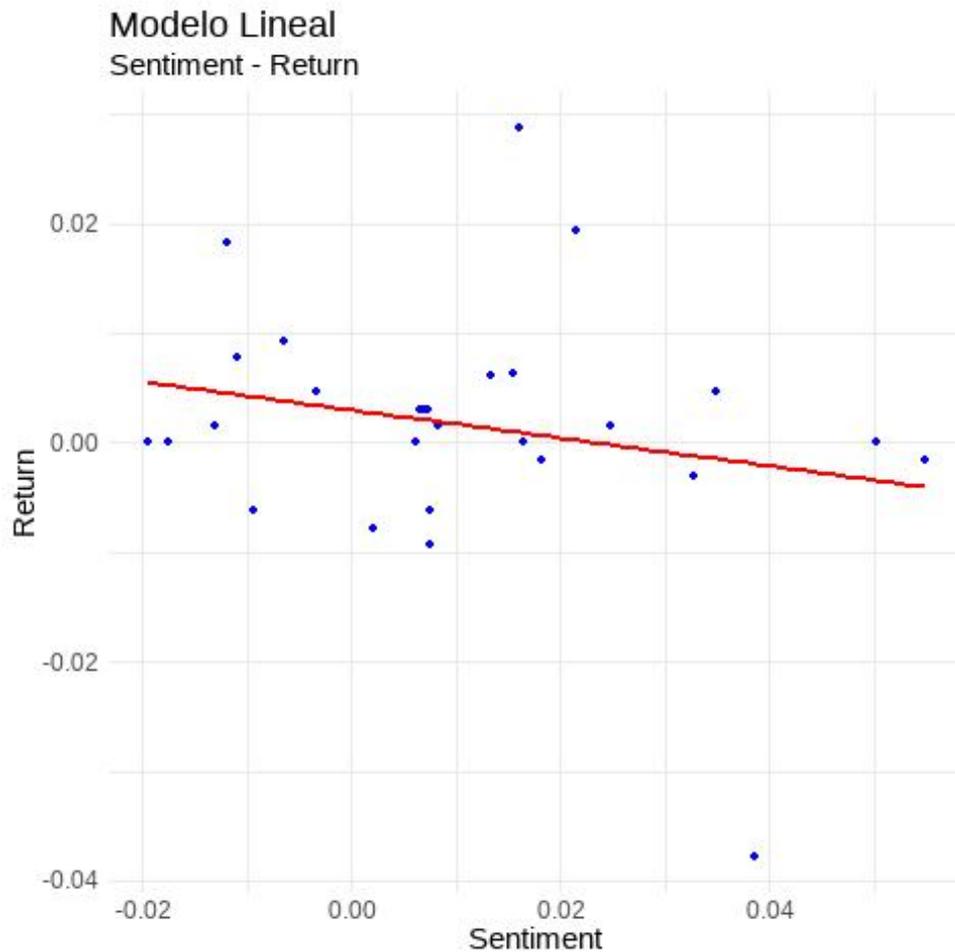


Figura 15. Modelo de Regresión Lineal Simple

Del gráfico anterior se observa que a menor sentimiento, es decir más cercano a cero, mayor es el retorno diario de la acción. Además, se observa que es factible trazar un modelo lineal de regresión entre ambas variables, teniendo en consideración al sentimiento como variable explicativa (o regresora) y como variable explicada (o respuesta) al rendimiento diario logaritmico de la acción local del Banco Santander Argentina.

Una limitación de este enfoque es que, si bien nos interesa cómo eventos que afectan negativamente a la reputación, pueden afectar a la valoración y a los futuros ingresos de la empresa, a veces la relación entre el sentimiento de *Twitter* y los retornos puede tener causalidad inversa. Un ejemplo de esto sería la noticia de que la cantidad de créditos otorgados por la compañía se desplomaron en el último trimestre, lo que previsiblemente haría que la cotización caiga, y a su vez, que los inversores comiencen a realizar comentarios negativos en *Twitter*. En este caso, no se estaría hablando de un evento que afecta la reputación de la empresa, lo que a su vez daña su desempeño económico. Sino de un mal desempeño económico que empeora la métrica de reputación obtenida de la red social. Este tipo de eventos pueden hacer que el capital en riesgo estimado con la técnica propuesta sea excesivamente alto, pues se estaría incluyendo eventos donde la relación causal no es la que nos preocupa. Desde este punto de vista, el enfoque aquí propuesto es un enfoque conservador. Para mitigar este problema, en el presente trabajo se utilizó una ventana temporal que no incluye fechas en las que la compañía presenta resultados. Esto, si bien no se puede asegurar que evite totalmente la problemática descrita, sin duda, minimiza

su posible impacto.

A continuación se presentan los resultados de las cuatro especificaciones del modelo econométrico utilizado para la estimación del coeficiente de interés. Los resultados se presentan en la siguiente tabla.

Tabla 5. Estimación de Especificaciones Propuestas

	<i>Variable Dependiente:</i>			
	<i>Stock</i>			
	(1)	(2)	(3)	(4)
<i>Market</i>	0.207* (0.118)		0.191* (0.092)	0.098 (0.202)
<i>Financiaci</i>				0.075 (0.363)
<i>Sentiment</i>		-0.128* (0.065)	-0.103* (0.056)	-0.105* (0.063)
<i>Constante</i>	0.001 (0.002)	0.003 (0.002)	0.002 (0.002)	0.002 (0.002)
Observaciones	28	28	28	28
R ²	0.502	0.582	0.632	0.656
Adjusted R ²	0.543	0.564	0.542	0.535
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01 ;		

Donde la variable *Market* representa el retorno diario del índice global *S&P Merval*. La variable *Financiaci* representa el retorno diario del índice sectorial *S&P Merval Financiero*, que sigue la evolución de las acciones vinculadas al sector financiero (bancario) que cotizan en el BYMA. Y la variable *Sentiment* representa el sentimiento de los *tweets* medido por la implementación del algoritmo de Feuerriegel *et al.* (2019) descrito en el Capítulo 2.

De la tabla anterior, se selecciona la especificación (2) del modelo por poseer el R^2 Ajustado más elevado. En términos formales,

$$Stock \hat{Return}_t = 0,003 - 0,128 * Sentiment_t$$

A partir de la especificación seleccionada se utilizó el coeficiente de regresión estimado para aproximarse el cálculo del Capital Económico por Riesgo Reputacional.

El valor de mercado V_t del Banco al 8 de Febrero de 2022 se determina como el producto de los siguientes valores,

1. Q_t : Cantidad de acciones en circulación al momento t
2. P_t : Precio de Cierre Ajustado de la Acción de BRIO.BA al momento t ³⁷

La cantidad de acciones en circulación son 4,397,500,060 según lo publicado en la página 6 de la última comunicación de la disciplina de mercado y el último precio de cierre ajustado se obtiene de información pública de mercado, que a la fecha de valuación fue de \$33,60 por acción³⁸. En términos formales,

$$V_t = Q_t * P_t = \frac{4,397,500,060}{1,000} * \$33,60 = \$147,756,002$$

³⁷Ticker del instrumento cotizante del Banco Santander Argentina S.A.

³⁸El informe de la disciplina de mercado se encuentra disponible en el siguiente link: bit.ly/3NzgdIN

Es decir, el valor de mercado al 8 de Febrero de 2022 es de \$147,756,002 en miles de Pesos Argentinos³⁹. Este monto estimado es lo que se denomina en la jerga financiera como *Market Cap*, el cual coincide con el valor publicado en *Yahoo Finance* y se contempla en la siguiente figura.



Figura 16. Información de mercado de la Acción BRIO.BA

Teniendo en cuenta la capitalización de mercado, el valor estimado del coeficiente de regresión al momento de la valuación, y el percentil 95 de la variable índice de reputación se calcula numéricamente el Capital Económico por Riesgo Reputacional al 8 de Febrero de 2022 mediante la ecuación (16),

$$Capital_{Rep} = \$147,756,002 * -0,127974 * -0,202263258 = \$427,957,80$$

Es decir, el capital en riesgo al 8 de Febrero de 2022 equivale a \$427,957,80 en miles de Pesos Argentinos y representa el 0,2896 % de la capitalización bursátil y un 0,2242 % de la Responsabilidad Patrimonial Computable (RPC)⁴⁰.

Para la organización este número debería ser indicativo del nivel de riesgo dispuesto a asumir por eventos reputacionales, los que podrían estar vinculados como por ejemplo con declaraciones públicas de quienes la dirigen o fallas en los servicios.

3.2.4. Sobre el Backtesting del modelo

Para realizar una validación del modelo estimado mediante un *Backtesting*, era necesario recolectar *tweets* para una ventana temporal más alejada en el tiempo y disponer de las cotizaciones de la acción del Banco Santander Argentina S.A. Por tal motivo, se recolectaron datos de opiniones públicas de *Twitter* desde el 21 de Marzo (incorporando información desde el 13 de Marzo) hasta el 5 de Mayo de 2022. Asimismo, se procedió de la misma manera que para la ventana de tiempo utilizada previamente. La base de *testing* está compuesta por aproximadamente 126.000 *tweets* únicos con 87 variables en 45 días de recolección.

³⁹Se publica en esta unidad de medida por una cuestión de practicidad.

⁴⁰A la fecha bajo análisis se utilizó el último dato de \$190,879,234 a Septiembre de 2021, disponible en el sitio web de la Entidad.

En el siguiente gráfico se exhibe la evolución diaria del indicador de sentimiento para la ventana utilizada en la prueba retrospectiva.

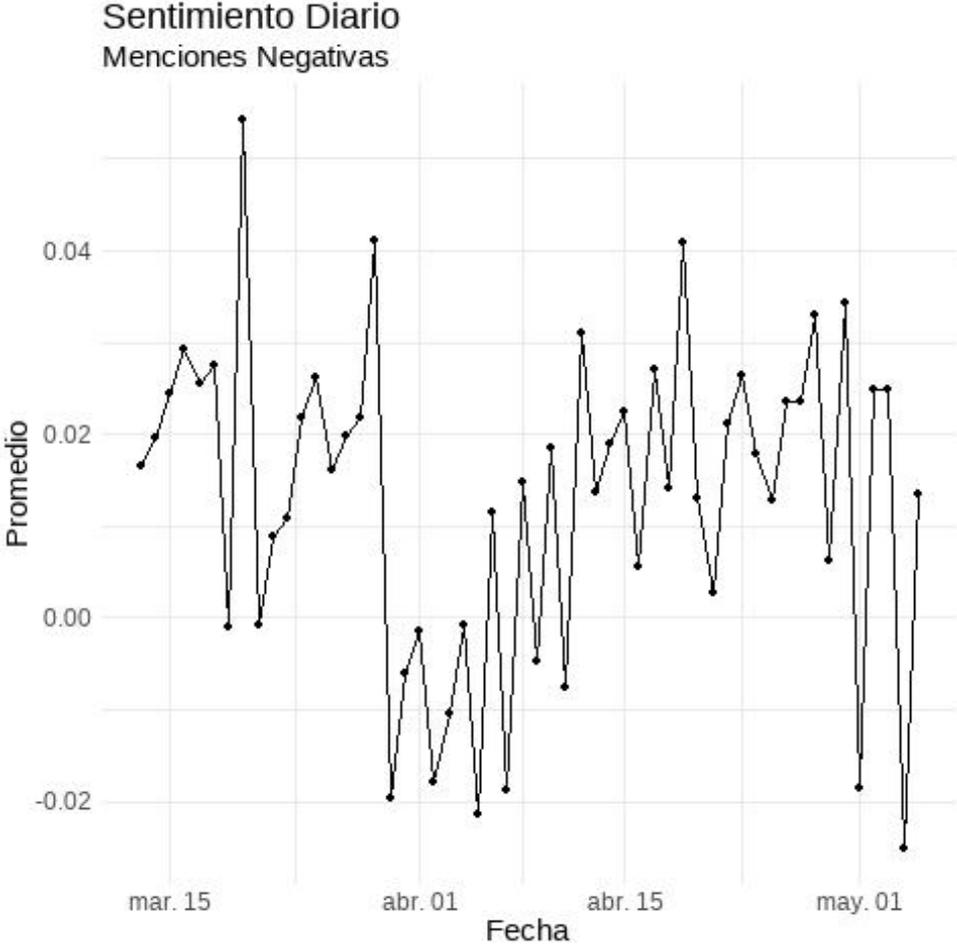


Figura 17. Sentimiento Diario de Menciones

Del gráfico anterior se observa que en la ventana analizada, la serie vinculada al indicador de reputación presentó una volatilidad mayor al período utilizado para estimar el Capital Económico. El desvío estándar del indicador es de 0,0321832031, mientras que en la ventana anterior fue de 0,0188660726. Al no contar con los precios de cotización del Banco en el período de recolección de datos para la prueba, no se ha podido validar el modelo empleado. En el siguiente gráfico se observa la cantidad de menciones por hora.

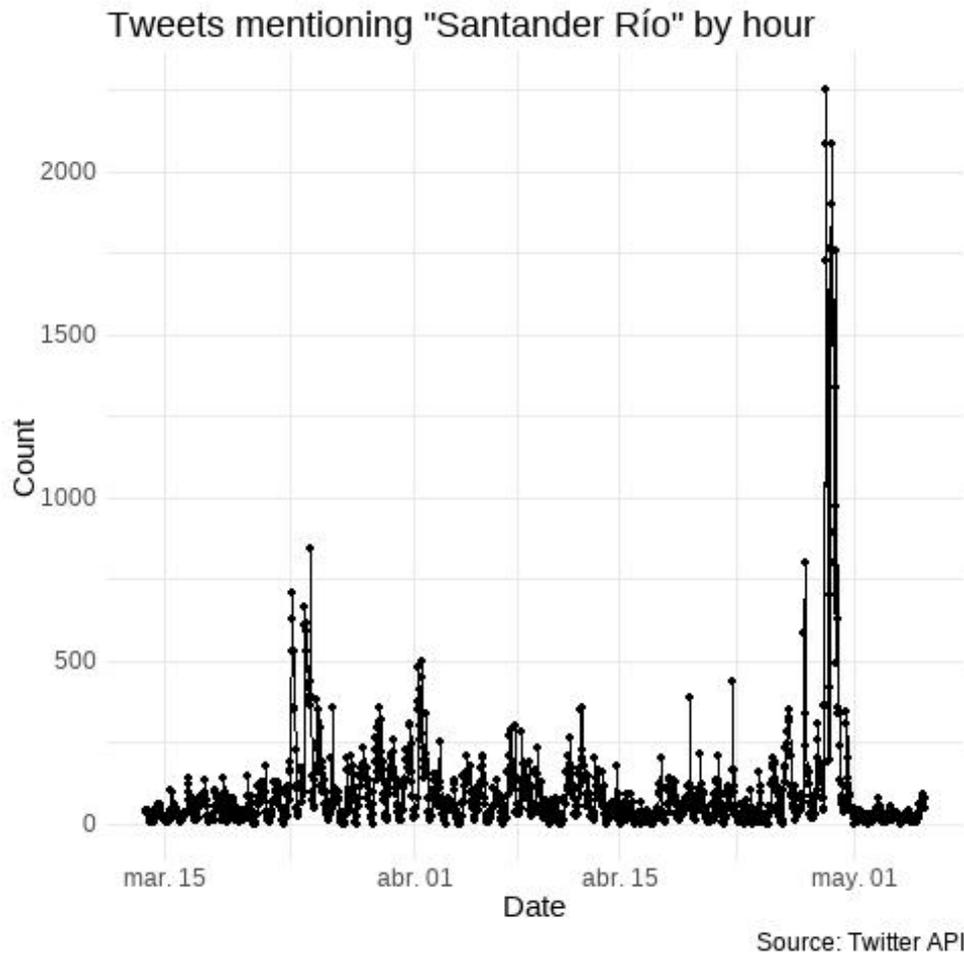


Figura 18. Cantidad de Menciones Negativas por Hora

En el siguiente gráfico se exhibe el histograma del indicador diario del sentimiento.

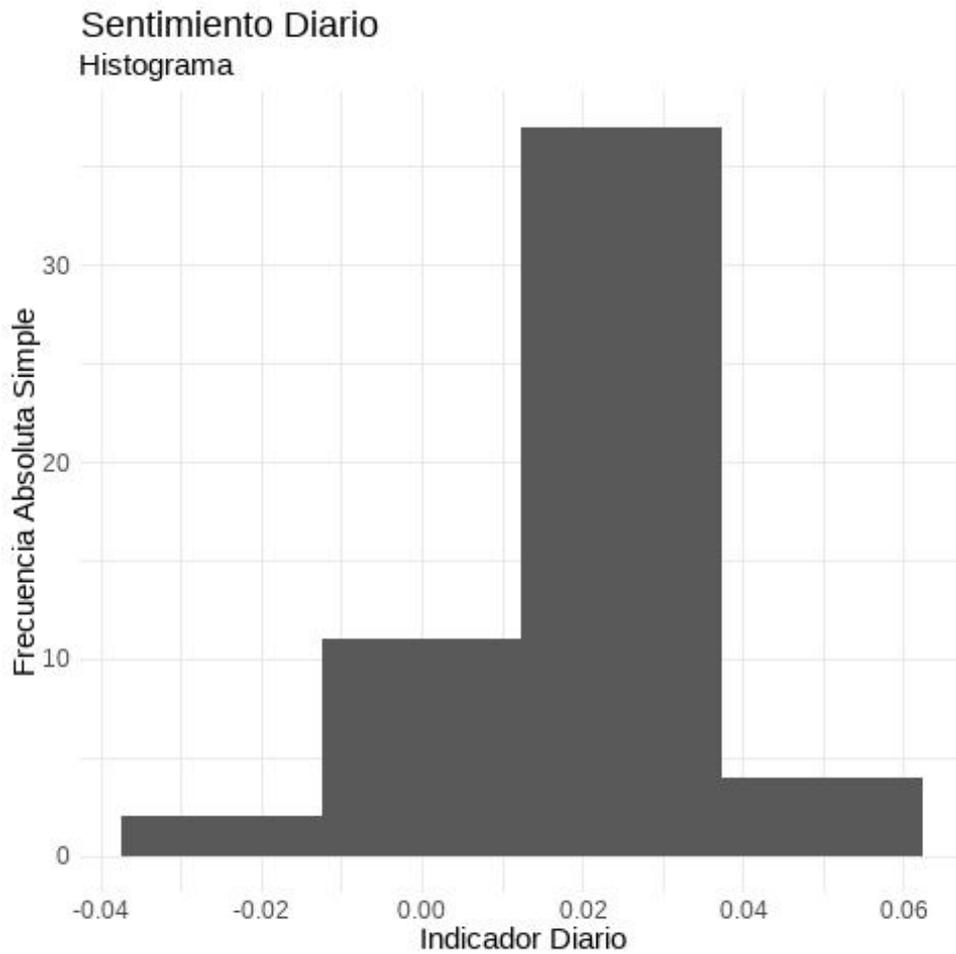


Figura 19. Histograma del Sentimiento Diario

La siguiente tabla resume medidas descriptivas sobre el indicador de reputación en la ventana seleccionada para realizar el *Backtesting*. En base a los valores obtenidos se observa que la variable posee asimetría negativa. Asimismo, el valor medio de este indicador es 0,0201371268 con un desvío de 0,0321832031 y el valor mediano es 0,0222332016.

Medida	Sentiment
Media	0.0201371268
Desvío	0.0321832031
Mediana	0.0222332016
Mínimo	-0.0238095238
Máximo	0.0416666667
Varianza	0.0002022132
Asimetría	-1.0411573978
Curtosis	1.0678947005
Percentil 5	-0.010416667
Percentil 95	0.038461538
n	54

Tabla 6. Medidas Descriptivas

3.3. Recomendaciones de Gestión

La propuesta de modelización presentada en este trabajo es una aproximación a la utilización de técnicas de *Text Mining* para la cuantificación de este riesgo y podría constituirse en una alternativa de medición para aquellas organizaciones que cotizan en algún mercado de valores. Además, se podría emplear como una herramienta para diseñar ciertos indicadores vinculados a la opinión pública que tengan los usuarios sobre la organización, y por lo tanto monitorear adecuadamente la reputación. Asimismo, el modelo estimado podría calibrarse diariamente a partir de la extracción de datos de *Twitter* y desarrollar un indicador diario de evolución del Capital Económico.

También es posible diseñar matrices de correlaciones entre los distintos riesgos, incluidos el Riesgo Reputacional a partir del indicador desarrollado. Además, podría emplearse para la realización de pruebas de estrés para determinar la resistencia de la organización ante escenarios adversos pocos probables pero posibles (plausibles) que puedan afectar a la empresa.

Tal como lo recomienda el Comité de Supervisión Bancaria de Basilea el Capital Económico debe ser suficiente para absorber las pérdidas inesperadas que enfrentan las Compañías Financieras. Por tal motivo, es preciso tener conocimiento del nivel de riesgo que asumen para proteger la solvencia del sistema financiero y minimizar el riesgo de quiebra o ruina.

Nótese que la incidencia de reservar capital para atender a este riesgo conlleva menos capacidad prestable que tendrá la empresa para otorgar créditos. No obstante, las ventajas de disponer de un monto destinado exclusivamente a situaciones adversas vinculadas a la reputación puede generar un beneficio a largo plazo porque la empresa destinaría un monto para cuantificar el impacto económico. En la situación de que se opte por gestionar este riesgo y reservar capital, se podría definir un proceso de validación del modelo mediante la realización de pruebas retrospectivas (*Backtesting*) para calibrar periódicamente los parámetros empleados. Además, la utilización de técnicas de *Analytics* permite adaptarse más rápidamente a los cambios que afectan al riesgo reputacional, dado que puede variar sustancialmente en el tiempo, de la misma forma que varían las percepciones y expectativas de los distintos *Stakeholders*.

Resumen de Capítulo

En el presente capítulo se realizó una aplicación práctica orientada a negocios, mediante la utilización de técnicas de *Text Mining*, en particular de *Sentiment Analysis*, para realizar una estimación del cálculo del Capital Económico que debería reservar el Banco Santander Argentina para hacer frente al Riesgo Reputacional.

Para cumplimentar con dicho objetivo se realizó un proceso de extracción de datos públicos de la red social *Twitter*, se resumió la información y se realizaron varias visualizaciones para entender que ocurrió. Luego se aplicaron técnicas de Análisis de Sentimiento para calcular el riesgo reputacional para el Banco Santander Argentina. Por último, se realizaron ciertas recomendaciones vinculadas a la gestión de este riesgo.

Esta página se deja en blanco intencionalmente

"...Y dale alegría, alegría a mi corazón, afuera se irán la pena y el dolor. Y ya verás, las sombras que aquí estuvieron ya no estarán..."

Fito Páez, *Y dale alegría a mi corazón*

Conclusiones y Futuras Investigaciones

El presente trabajo constituye una aproximación a la medición del Riesgo Reputacional de organizaciones y en particular de Entidades Financieras que cotizan en un mercado de valores. Tal como se ha expresado previamente, constituye un riesgo difícil de medir y cuantificar dentro del Proceso de Gestión del Riesgo que implementan los bancos según lo recomendado por el Comité de Supervisión Bancaria de Basilea.

El objetivo general de este trabajo es estimar el cálculo del Capital Económico por Riesgo Reputacional mediante la utilización de técnicas de *Text Mining*. En esta tesis se dio respuesta y se desarrolló una metodología propia para medir el efecto del riesgo reputacional en términos del riesgo de mercado (precio), tal como lo establece el Banco Central de la República Argentina.

El modelo se desarrolló mediante la utilización de *Sentiment Analysis* sobre datos públicos extraídos de las opiniones de usuarios de *Twitter* y la obtención del precio de cotización del Banco Santander Argentina S.A. en el mercado BYMA.

Con respecto a la limitación del presente trabajo, puede mencionarse que no se pudo validar el modelo mediante un *Backtesting* porque el Banco analizado dejó de cotizar en el período de prueba.

Luego, existen ciertos inconvenientes operativos que no representan limitaciones del trabajo. Estos problemas han sido resueltos, no obstante si se contase con otros recursos informáticos como más memoria RAM y la credencial paga para acceder a la API de *Twitter* permitirían mejorar el análisis realizado. Por ejemplo, la utilización de la API gratuita de *Twitter* únicamente permite extraer *tweets* de los últimos 10 días corridos. No obstante, esto se subsanó mediante la recolección diaria durante un mes y su almacenamiento en archivos en forma json para su posterior análisis. Además, se observó al momento de procesar los datos de la base total, una imposibilidad para calcular el análisis de sentimiento de tantas palabras por el procesamiento de los datos.

Para futuras mejoras del modelo exhibido se propone incluir a los resultados del sentimiento de las redes sociales, otras noticias económicas a través de publicaciones en diarios y revistas, como así también análisis de series temporales vinculadas a variables económicas y financieras, como por ejemplo la inclusión de la importancia de las menciones de *Google Trends* mediante la API disponible. En este trabajo se utiliza un modelo lineal de regresión, sin embargo podría emplearse una metodología de análisis de series temporales para predecir el indicador de reputación negativa⁴¹.

⁴¹Tales como modelos ARIMA para análisis univariados o modelos VAR para multivariados, entre otras.

Además, este modelo podría replicarse para otra entidad financiera y validarse sus resultados con una prueba retrospectiva sobre la estimación del Capital Económico para determinar si el nivel reservado es adecuado a los riesgos asumidos. Para tal fin, sería necesario relevar más datos y tener como mínimo un año de estimación del Capital Económico según las recomendaciones del Comité de Supervisión Bancaria de Basilea. Otra cuestión que podría incluirse en una futura investigación es la realización de Pruebas de Estrés individual sobre el Riesgo Reputacional de la organización, como así también el desarrollo de un Plan de Contingencia. Por último, tener presente que esta metodología podría mejorarse mediante la incorporación de otros factores vinculados a otros riesgos correlacionados con el reputacional, como por ejemplo el riesgo de liquidez, riesgo de tasa de interés, riesgo de crédito, entre otros. Asimismo, como se mencionó en el apartado de recomendaciones de gestión, es posible diseñar matrices de correlaciones entre los distintos riesgos a los que se expone la entidad con la finalidad de dimensionar el grado de asociación de los mismos.

"...Todo al fin se sucedió, solo que el tiempo no los esperó, la melancolía de morir en este mundo y de vivir sin una estúpida razón..."

Fito Páez, *Mariposa tecknicolor*

Referencias Bibliográficas

Andersen, S., Fountain, J., Harrison, G. (2014). Estimating subjective probabilities. *J Risk Uncertain* 48, 207–229. <https://doi.org/10.1007/s11166-014-9194-z>.

Arias, M., Arratia, A., Xuriguera, R. (2012). Forecasting with Twitter Data. *Universitat Politècnica de Catalunya*. <https://dl.acm.org/doi/10.1145/2542182.2542190>

Aromí, D. (2022). Medición de Incertidumbre Económica en Redes Sociales en Base a Modelos de Procesamiento de Lenguaje Natural. Serie Documentos de Trabajo N° 179, Red Nacional de Investigadores de Economía. <https://rednie.eco.unc.edu.ar/files/DT/179.pdf>

Aromí, D., Llada, M (2022). Forecasting inflation with Twitter. Serie Documentos de Trabajo del IIEP, 76, 1-20. <https://ojs.econ.uba.ar/index.php/DT-IIEP/article/view/2620>

Basel Committee on Banking Supervision. (2004). Principles for the Management and Supervision of Interest Rate Risk. Recuperado de *BCBS*

Basel Committee on Banking Supervision (2016). Interest rate in the banking book. Standards. Bank for International Settlements. Recuperado de *BCBS*

Bessis, J. (2002). Risk Management in Banking. Second Edition. John Wiley & Sons Ltd, England.

Dave, K., Lawrence, S., Pennock, D. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In Proceedings of International Conference on World Wide Web (WWW-2003).

Deveikyte, J., Geman, H., Piccari, C., Proveti, A. (2020). A Sentiment Analysis Approach to the Prediction of Market Volatility. Birkbeck, University of London.

Feuerriegel, S., Pröllochs, N., Neumann, D. (2019). Statistical inferences for polarity identification in natural language. *PLoS ONE* 13(12): e0209323. <https://doi.org/10.1371/journal.pone.0209323>.

Fiordelisi, F., Schwizer, P., Soana, M. (2011). The Determinants of Reputational Risk in the Banking Sector. *Journal of Banking and Finance*. Recuperado de *DOI: 10.2139/ssrn.1895327*

Freixas, X., Rochet, J. (1999). *Economía Bancaria*. Antoni Bosch. España.

Gillet, R., Hubner, G., Plunus, S. (2010). Operational risk and reputation in the financial industry. *Journal of Banking & Finance*. Elsevier, ISSN 0378-4266. - Vol. 34.2010, 1, p. 224-235.

James, G., Hastie, T., Tibshirani, R., Witten, D. (2014). *An Introduction to Statistical Learning with Applications in R*. Springer Texts in Statistics. Second Edition.

Jiménez-Zafra, S., Martín-Valdivia, T., Ureña-López, A. (2020). Nuevos retos en el Análisis de Sentimientos. Programa de Doctorado “Tecnologías de la Información y la Comunicación”. Departamento de Informática. Universidad de Jaén. España.

Jorion, P. (2011). *Value at Risk: The New Benchmark for Managing Financial Risk*. Third Edition. McGraw-Hill.

Kordonis, J., Arampatzis, A., Symeonidis, S. (2016). Stock Price Forecasting via Sentiment Analysis on Twitter. Electrical and Computer Engineering Dept. Democritus University of Thrace. Greece.

Kukich, K. (1992). Techniques for automatically correcting words in text. *ACM Computing Surveys*, 24(4):377-439.

Kukk, K. (2018). Correlation between emotional tweets and stock prices. Bachelor's Thesis. Department of Linguistics and Philology. Uppsala University.

Kumar, V., Singla, S., Shalika, K., Chadha R. (2023). Stock Market Forecasting Using Sentiment Analysis and Deep Learning. *Advances in Information Communication Technology and Computing*. 10.1007/978-981-19-9888-1_48. (587-595). https://link.springer.com/10.1007/978-981-19-9888-1_48

Kwartler, T. (2017). *Text Mining in Practice with R*, First Edition. John Wiley & Sons Ltd.

Laveglia, E., Garcia Fronti, J. (2018). El impacto de los medios sociales en la gestión del riesgo reputacional de las marcas. *Revista de Dirección y Administración de Empresas*. Número 25, págs. 75-94.

Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments and Emotions*. Cambridge University Press. United States of America.

Loughran, McDonald (2011). When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *Journal of Finance*, 66:1, 35-65.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J. and Barton, D. (2012) Big Data. The Management Revolution. *Harvard Business Review*, 90, 61-67.

Mitic, P. (2008). Reputation Risk: Measured. *International Journal of Safety and Security Engineering*. Vol. 8, No. 1, 171-180.

Munafo, F. (2019). Implementación de Text Mining en R para predecir el riesgo de default corporativo. (Trabajo Final de Posgrado. Universidad de Buenos Aires.). Recuperado de [Link](#)

- Nasukawa, T., Yi, J. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In Proceedings of the K-CAP-03, 2nd International Conference on Knowledge Capture.
- Novello, D. (2008). El Riesgo de Crédito en Basilea II. EDICON. Fondo Editorial Consejo. 1era Edición. Ciudad Autónoma de Buenos Aires. Argentina.
- Perales González, C., Marrao Rodrigues, H., Rodríguez-Oliveros, R. (2017). Sentiment analysis as reputational risk indicator. Disponible en SSRN: <http://dx.doi.org/10.2139/ssrn.3051870>.
- Schmarzo, B. (2013). Big Data: Understanding How Data Powers Big Business. Wiley. United States of America.
- Soprano, A., Crielaard, B., Piacenza, F., Ruspantini, D.(2009). Measuring Operational and Reputational Risk. A Practitioner´s Approach. John Wiley & Sons Ltd.
- Turney, P., Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. Journal of Artificial Intelligence Research, 37(1):141–188.
- Vadivukarassi, M., Puviarasan, N., Aruna, P. (2017). Sentimental Analysis of Tweets Using Naive Bayes Algorithm. World Applied Sciences Journal 35 (1): 54-59. ISSN 1818-4952.
- Vidya, N., Fanany, M., Budi, I. (2015). Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers. Procedia Computer Science 72. 519 – 526
- Villanueva, J. (2020). El riesgo reputacional es ya un riesgo financiero. Recuperado de *Guía de Financiación Empresarial*
- Weiss, S., Indurkha, N., Zhang, T., Damerou, F. (2005). Text Mining: Predictive Methods for Analyzing Unstructured Information. Springer New York, New York, NY.

"...Nunca me creo en la cima o en la gloria, eso es un gran fantasma..."

León Gieco, *La colina de la vida*

Bibliografía

- Abbass, H. (2015). Computational Red Teaming. Risk Analytics of Big-Data-to-Decisions Intelligent Systems. Springer.
- Beltrán, H. (2021). Modelización del cálculo de la Exigencia de Capital Económico por Riesgo Reputacional. (Trabajo Final de Posgrado. Universidad de Buenos Aires.)
- Christodoulakis, G., Satchell, S. (2008). The Analytics of Risk Model Validation. Quantitative Finance Series.
- Consolandi, C., Jaiswal-Dale, A., Gabbi, G. (2009). US Financial Institutions: Reputational Risk and Senior Management Sell Decisions.
- Danneman, N., Heimann, R. (2014). Social Media Mining with R. Packt Publishing Ltd.
- Das, S., Chen, M. (2001). Yahoo! For Amazon: Extracting Market Sentiment from Stock Message Boards. In Proceedings of APFA-2001.
- Dean, J. (2014). Big Data, Data Mining and Machine Learning: Value Creation for Business Leaders and Practitioners. Wiley, First Edition, United States of America.
- Douglas, L. (2011). 3d Data Management: Controlling data volume, velocity and variety. Gartner. Retrieved.
- Elkan, C. (2001). Method and System for Selecting Documents by Measuring Document Quality.
- Kushal, D., Lawrence, S., Pennock, D. (2003). Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In Proceedings of International Conference on World Wide Web (WWW-2003).
- Linoff, G., Berry, M. (2011). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley Publishing, Inc., Indianapolis, Indiana.
- Morinaga, S., Yamanishi, K., Tateishi, K., Fukushima, T. . (2002). Mining Product Reputations on the Web. In Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2002).
- Pang, B., Lee, L., Vaithyanathan, S. (2002). Thumbs Up? Sentiment Classification Using Machine Learning Techniques. In Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2002).
- Panjer, H. (2006). Operational Risk. Modeling Analytics. John Wiley & Sons Ltd.

"...Miren todos, ellos solos, pueden más que el amor y son más fuertes que el Olimpo..."

Fito Páez, *11 y 6*

Apéndices

Datos

En el siguiente [link](#) de Google Drive se acceden a los archivos json recolectados diariamente mediante la rutina que figura en el apéndice de rutinas **Recolección de Tweets**.

Códigos

En el siguiente [link](#) de Github se disponibilizan los códigos utilizados en el presente trabajo.

Tablas de Formas Funcionales

La siguiente tabla extraída del libro de Zizka *et al.* (2020, pp. 111) exhibe distintas formas funcionales de cálculo del peso local,

Weight name	Calculation
Binary (Term Presence)	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = 1$ if $f_{ij} > 0$
Term Frequency (TF)	$lw_{ij} = f_{ij}$
Squared TF	$lw_{ij} = f_{ij}^2$
Thresholded TF	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = 1$ if $f_{ij} = 1$ $lw_{ij} = 2$ if $f_{ij} \geq 2$
Logarithm	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = \log(f_{ij} + 1)$ if $f_{ij} > 0$
Alternate Logarithm	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = 1 + \log f_{ij}$ if $f_{ij} > 0$
Normalized Logarithm	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = \frac{1 + \log f_{ij}}{1 + \log a_j}$ if $f_{ij} > 0$
Augmented Normalized TF	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = k + (1 - k) \left(\frac{f_{ij}}{x_j}\right)$ if $f_{ij} > 0$
Changed-coefficient Average TF	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = k + (1 - k) \frac{f_{ij}}{x_j}$ if $f_{ij} > 0$
Square Root	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = \sqrt{f_{ij} - 0.5} + 1$ if $f_{ij} > 0$
Augmented Logarithm	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = k + (1 - k) \log(f_{ij} + 1)$ if $f_{ij} > 0$
Augmented Average TF	$lw_{ij} = 0$ if $f_{ij} = 0$ $lw_{ij} = k + (1 - k) \frac{f_{ij}}{a_j}$ if $f_{ij} > 0$
DFR-like Normalization	$lw_{ij} = f_{ij} * \frac{a_j}{T_j}$
Okapi's TF Factor	$lw_{ij} = \frac{f_{ij}}{2 + f_{ij}}$

Figura 20. Formas Funcionales del Peso Local

La siguiente tabla extraída del libro de Zizka *et al.* (2020, pp. 113) exhibe distintas formas funcionales de cálculo de la ponderación global,

<i>Weight name</i>	<i>Calculation</i>
None	$gw_i = 1$
Inverse Document Frequency (IDF)	$gw_i = \log \frac{N}{n_i}$
Squared IDF	$gw_i = \log^2 \frac{N}{n_i}$
Probabilistic IDF	$gw_i = \log \frac{N-n_i}{n_i}$
Global frequency IDF	$gw_i = \frac{f_i}{n_i}$
Entropy	$gw_i = 1 + \sum_{j=1}^N \frac{f_j}{f_i} \log \frac{f_j}{f_i} / \log N$,
Incremented global frequency IDF	$gw_i = \frac{f_i}{n_i} + 1$
Log-global frequency IDF	$gw_i = \log \left(\frac{f_i}{n_i} + 1 \right)$
Square root global frequency IDF	$gw_i = \sqrt{\frac{f_i}{n_i} - 0.9}$
Inverse total term frequency	$gw_i = \log \frac{\sum_{i=1}^N t_j}{f_i}$

Figura 21. Formas Funcionales del Peso Global

La siguiente tabla extraída del libro de Zizka *et al.* (2020, pp. 113) exhibe distintas formas funcionales del factor de normalización,

<i>Weight name</i>	<i>Calculation</i>
None	$n_j = 1$
Cosine	$n_j = \sqrt{\sum_{i=1}^m (gw_i * lw_{ij})^2}$
Sum of weights	$n_j = \sum_{i=1}^m gw_i * lw_{ij}$
Max weight	$n_j = \max gw_i * lw_{ij}$
Max TF	$n_j = 0.5 + 0.5 * \frac{gw_i * lw_{ij}}{\max gw_i * lw_{ij}}$
Square root	$f_{ij} = \sqrt{gw_i * lw_{ij}}$
Logarithm	$f_{ij} = \log gw_i * lw_{ij}$
Fourth normalization	$n_j = \sum_{i=1}^m (gw_i * lw_{ij})^4$

Figura 22. Formas Funcionales del Factor de Normalización

Gráficos

Análisis de Normalidad sobre los retornos diarios

Normality Diagnosis Plot (hipo)

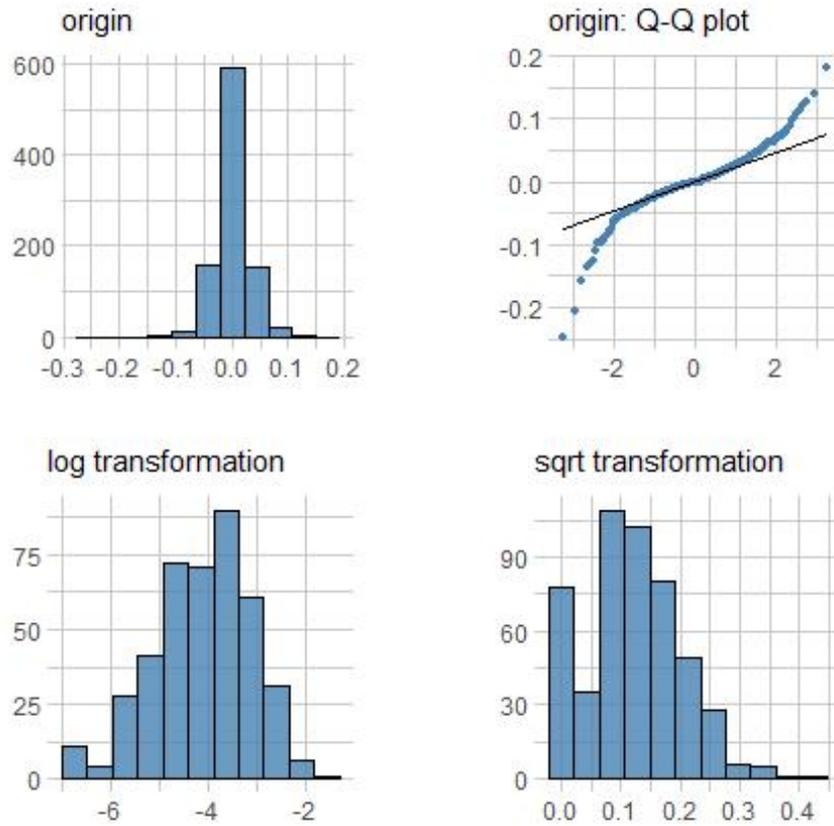


Figura 23. Diagnóstico de Normalidad - Retornos Hipotecario

Normality Diagnosis Plot (valo)

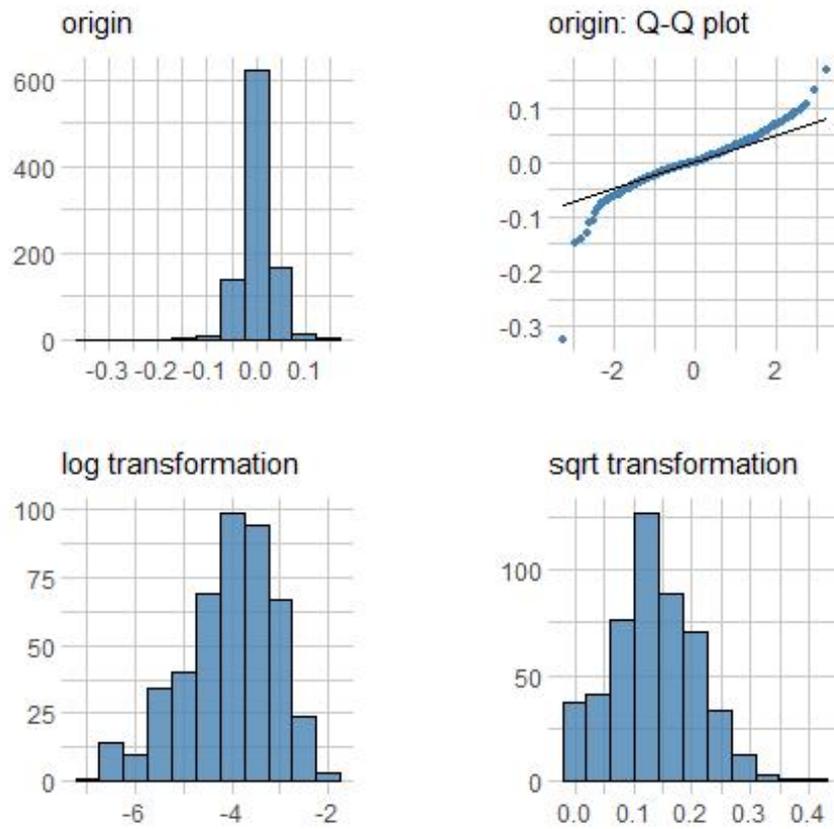


Figura 24. Diagnóstico de Normalidad - Retornos Valores

Normality Diagnosis Plot (macro)

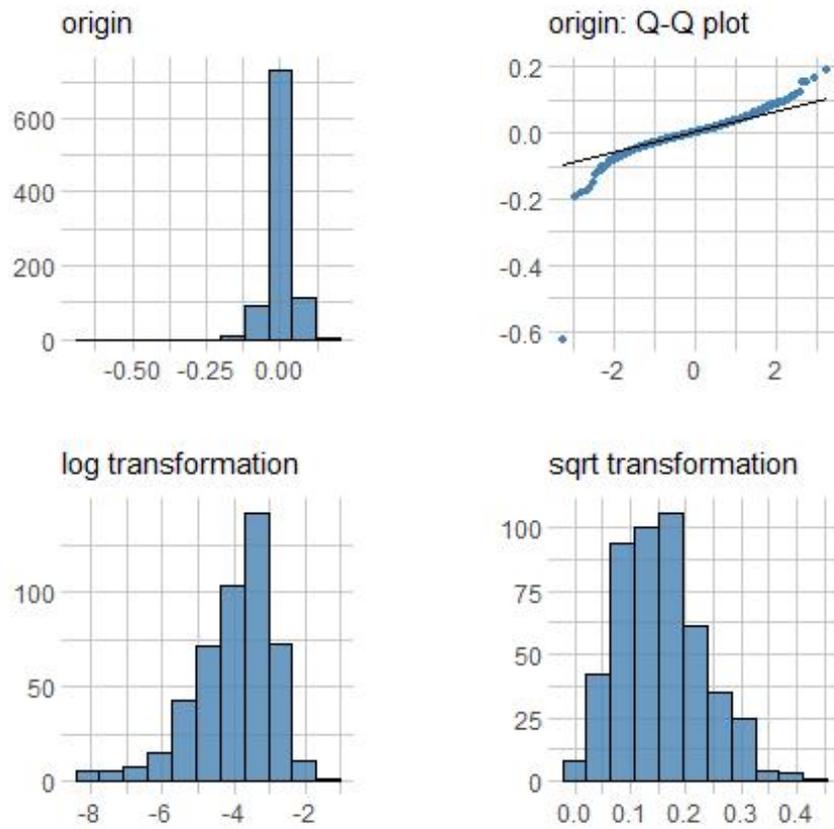


Figura 25. Diagnóstico de Normalidad - Retornos Macro

Normality Diagnosis Plot (pata)

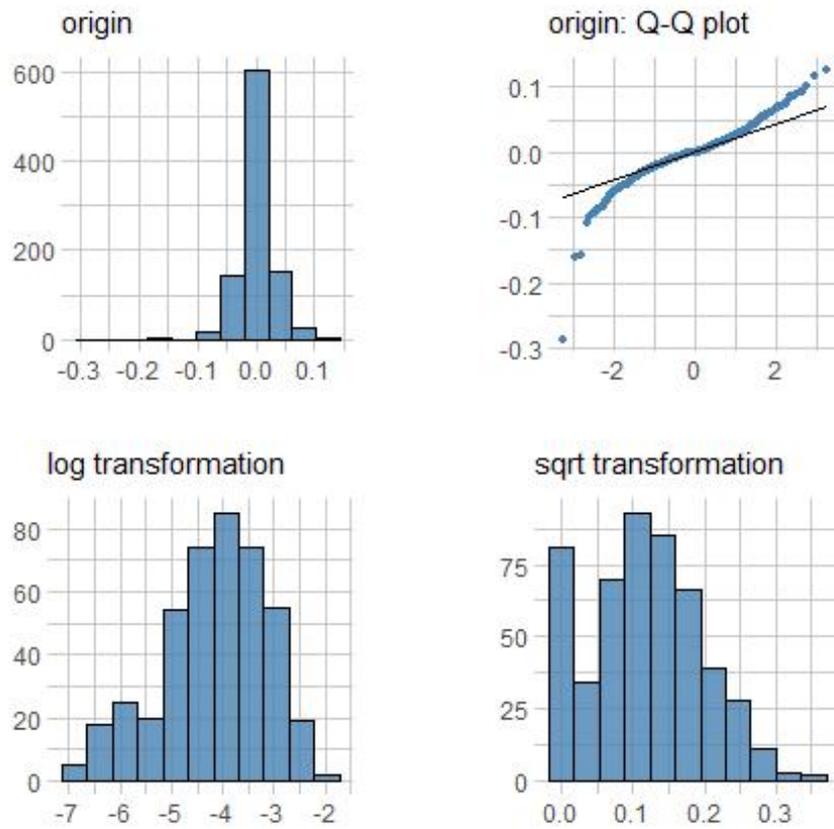


Figura 26. Diagnóstico de Normalidad - Retornos Patagonia

Normality Diagnosis Plot (bbar)

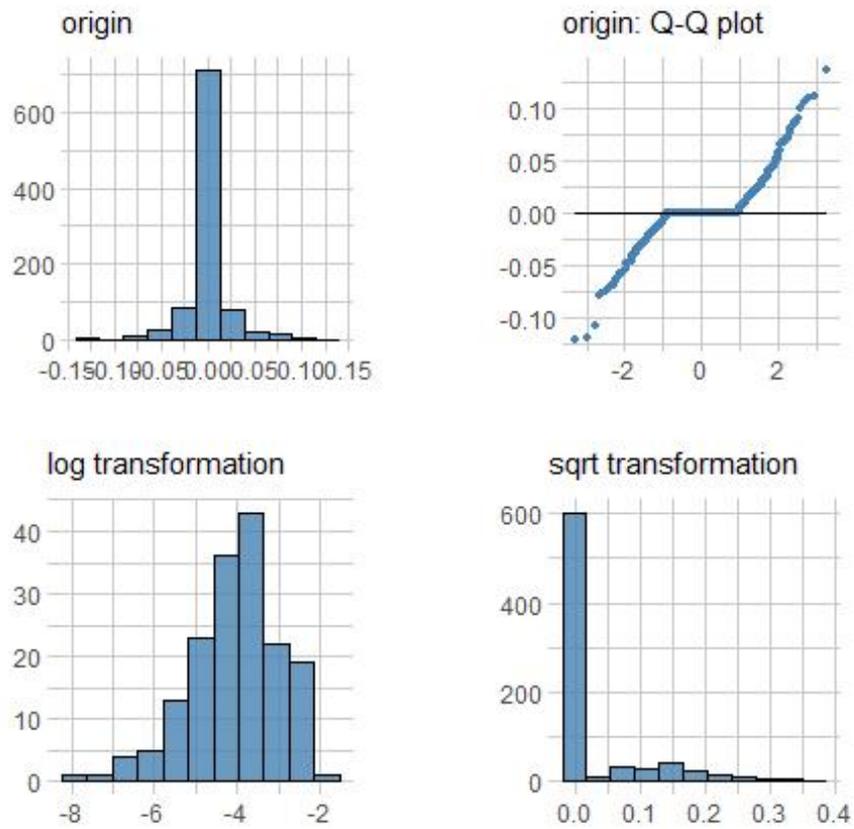


Figura 27. Diagnóstico de Normalidad - Retornos BBVA Francés

Normality Diagnosis Plot (brio)

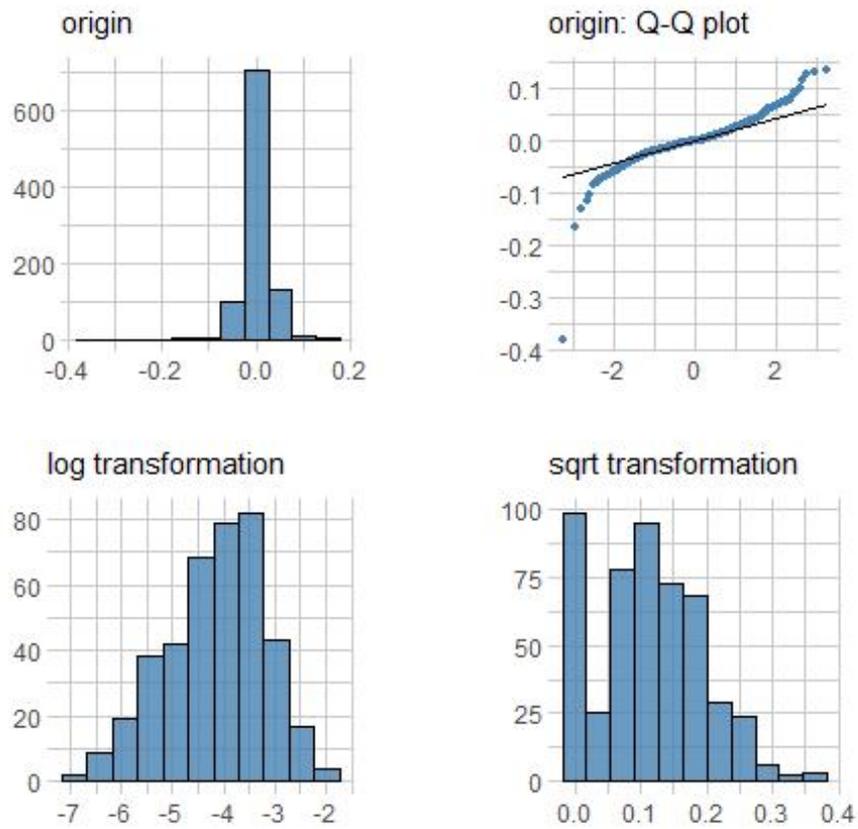


Figura 28. Diagnóstico de Normalidad - Retornos Santander Argentina

Normality Diagnosis Plot (ggal)

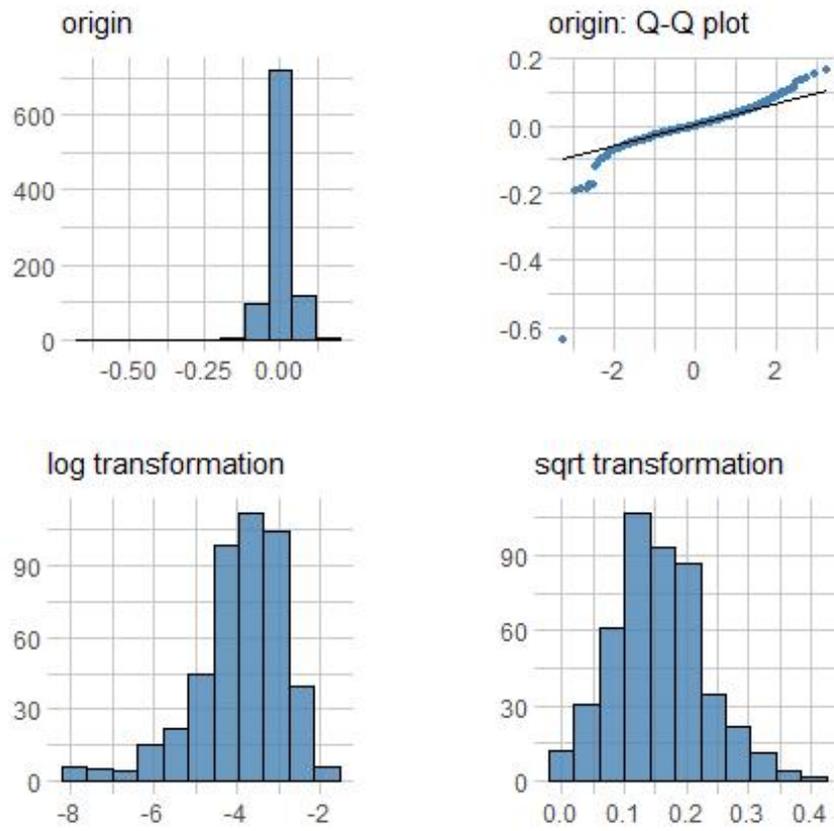


Figura 29. Diagnóstico de Normalidad - Retornos Galicia

Correlación de Retornos Diarios

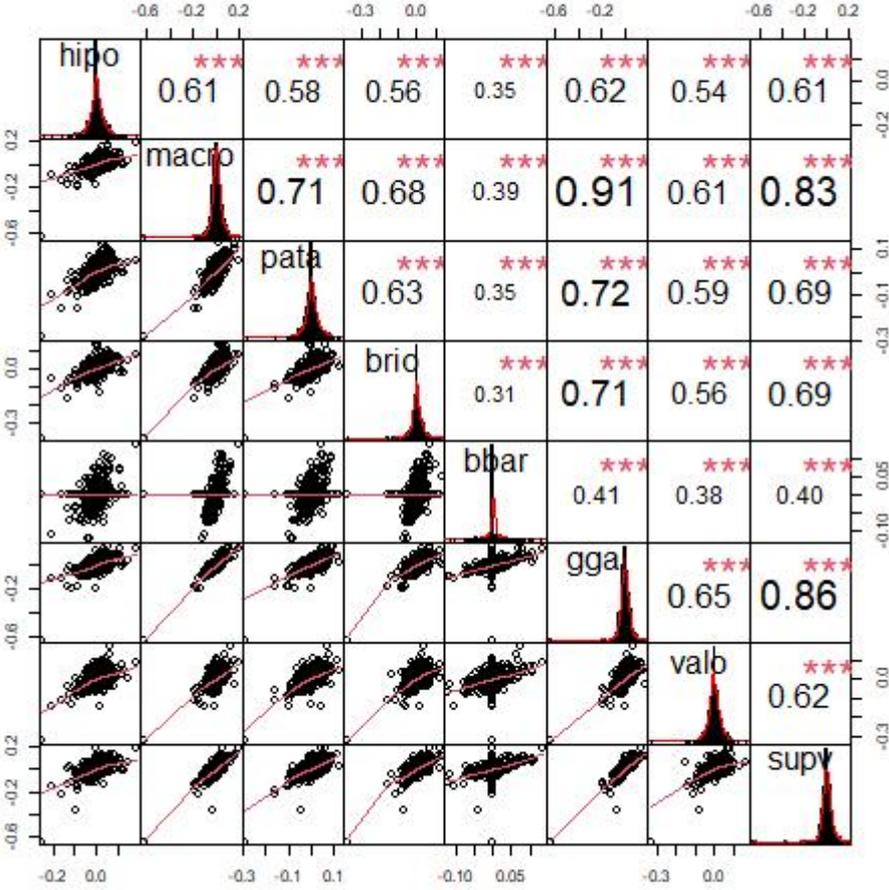


Figura 30. Correlación - Retornos Diarios

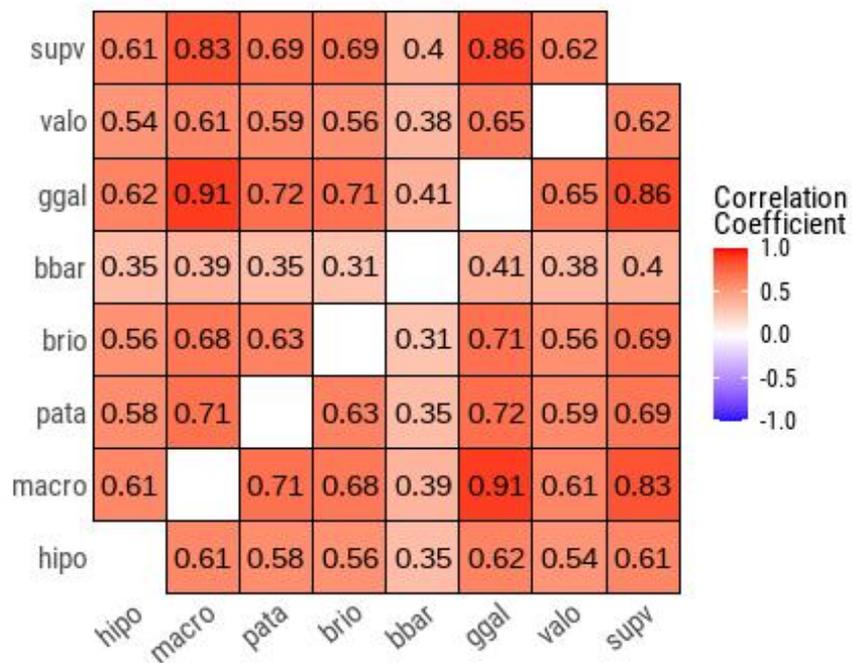


Figura 31. Correlación - Retornos Diarios

Análisis de Outliers de Retornos Diarios

Outlier Diagnosis Plot (hipo)

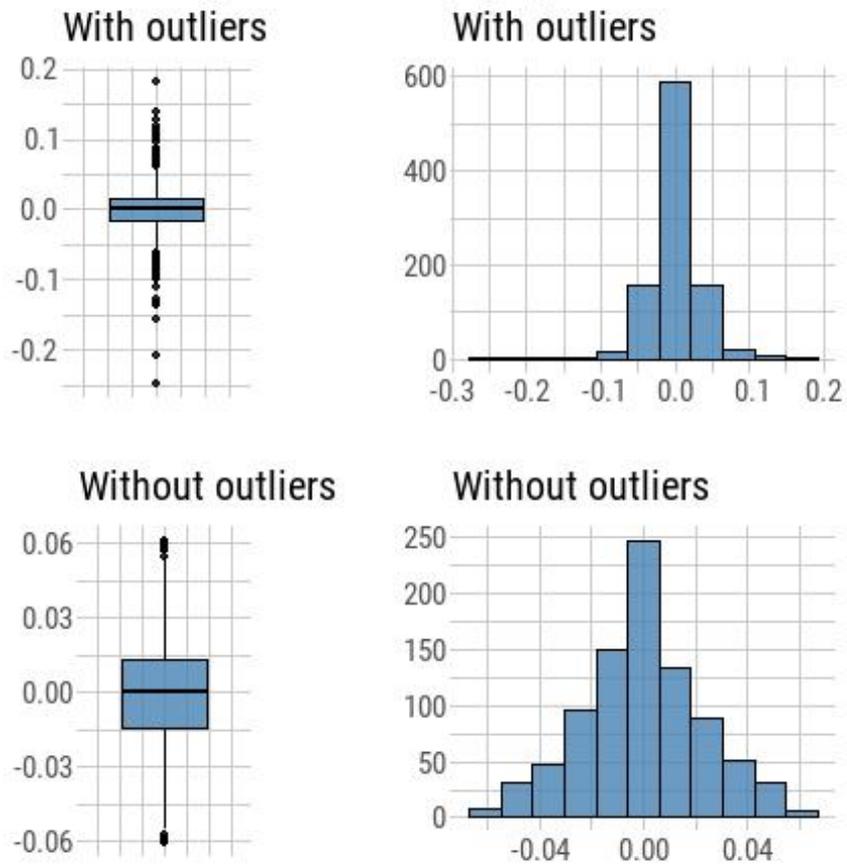


Figura 32. Outliers - Banco Hipotecario

Outlier Diagnosis Plot (valo)

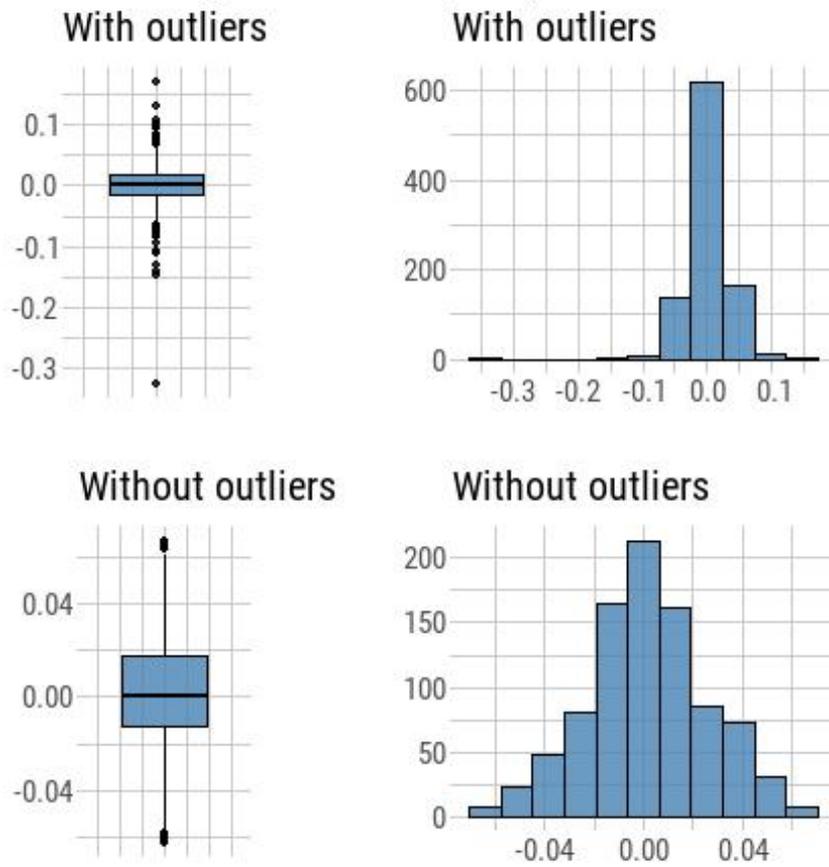


Figura 33. Outliers - Banco de Valores

Outlier Diagnosis Plot (macro)

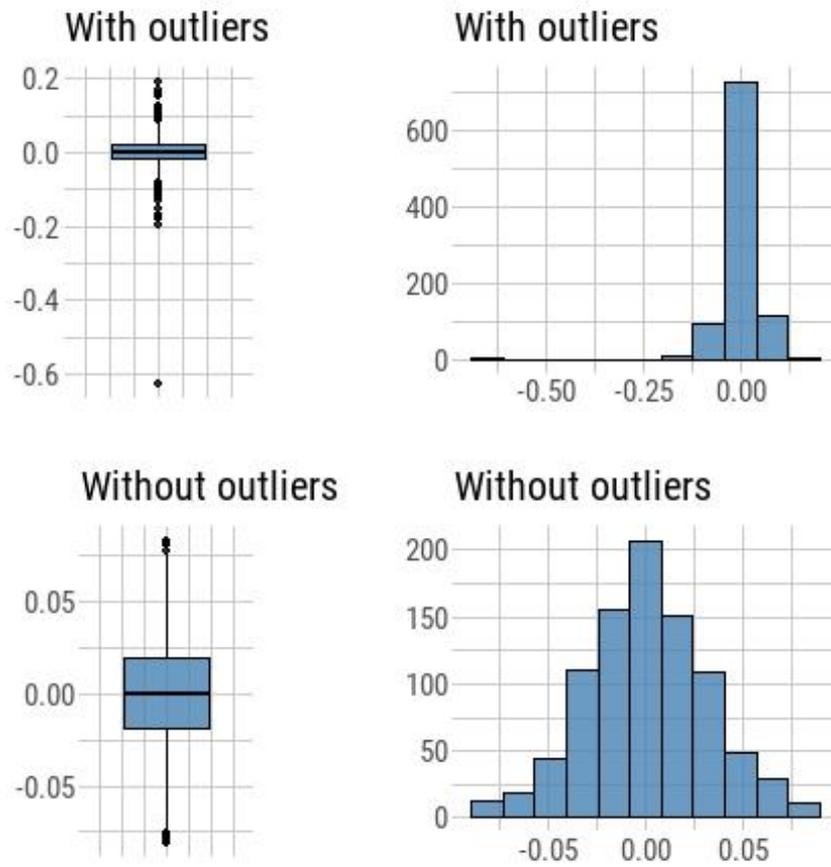


Figura 34. Outliers - Banco Macro

Outlier Diagnosis Plot (pata)

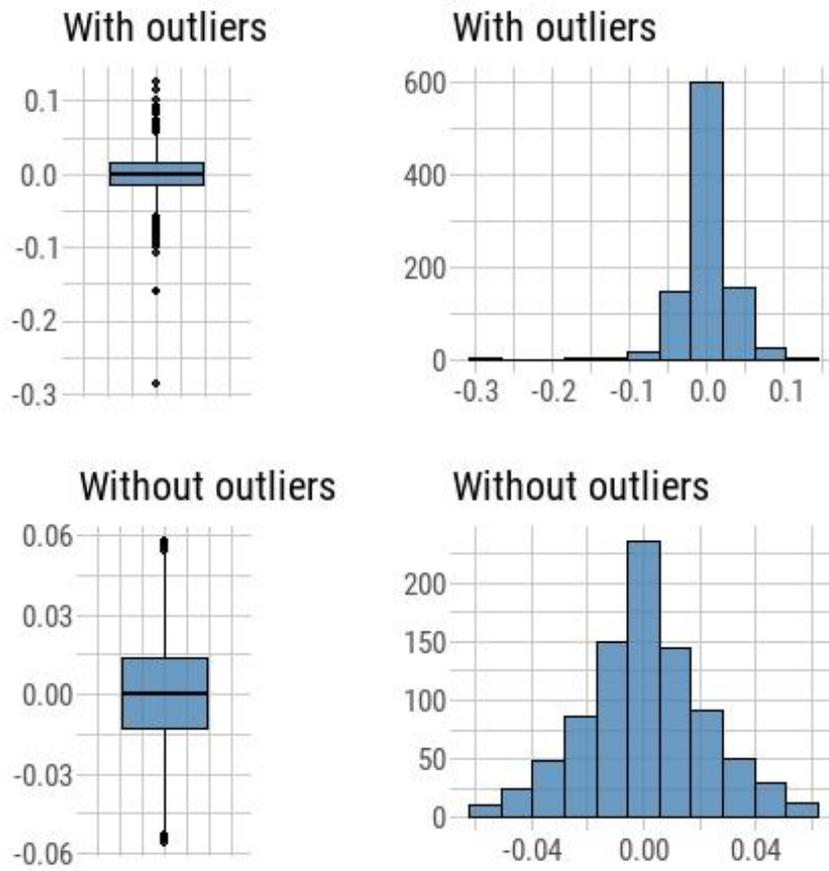


Figura 35. Outliers - Banco Patagonia

Outlier Diagnosis Plot (bbar)

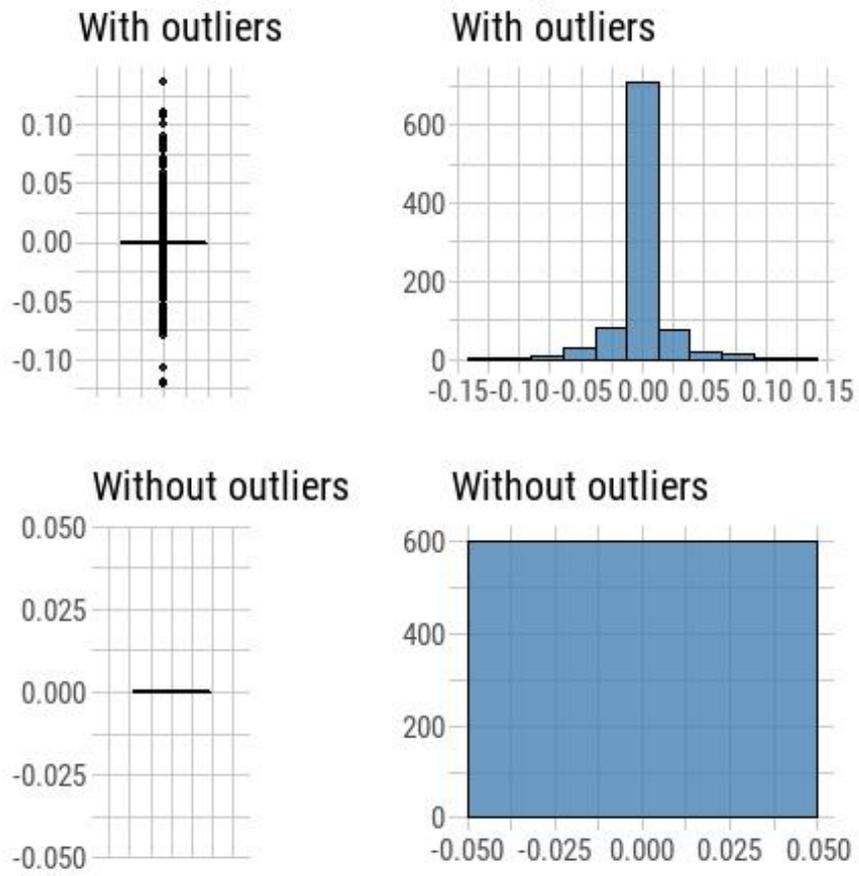


Figura 36. Outliers - Banco BBVA Francés

Outlier Diagnosis Plot (brio)

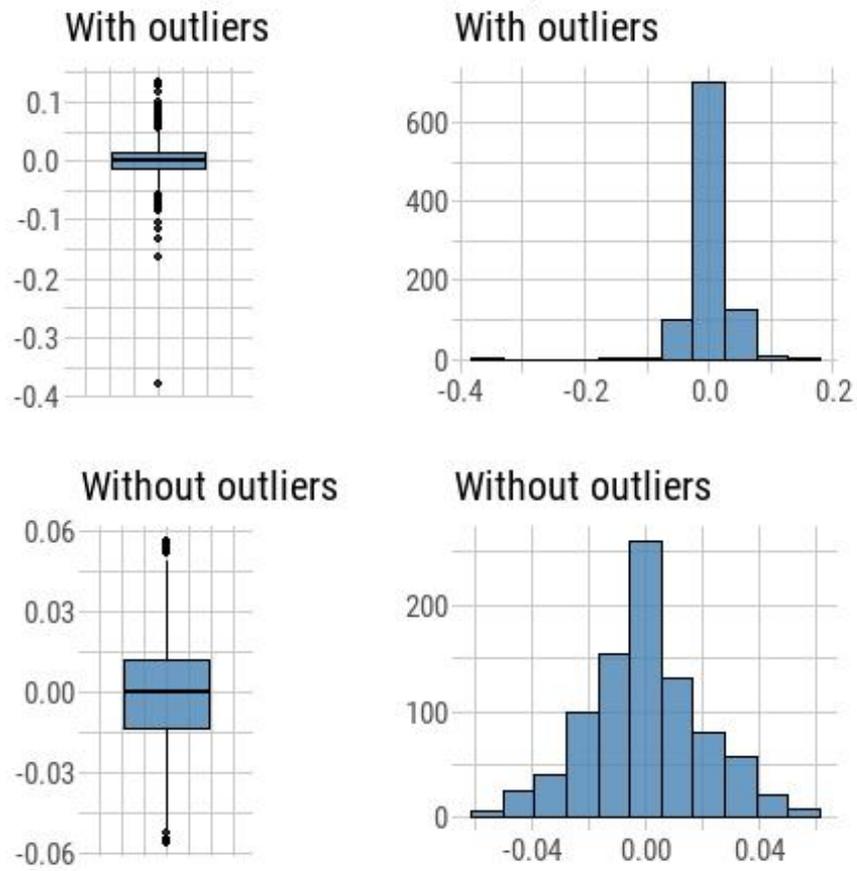


Figura 37. Outliers - Banco Santander Argentina

Outlier Diagnosis Plot (ggal)

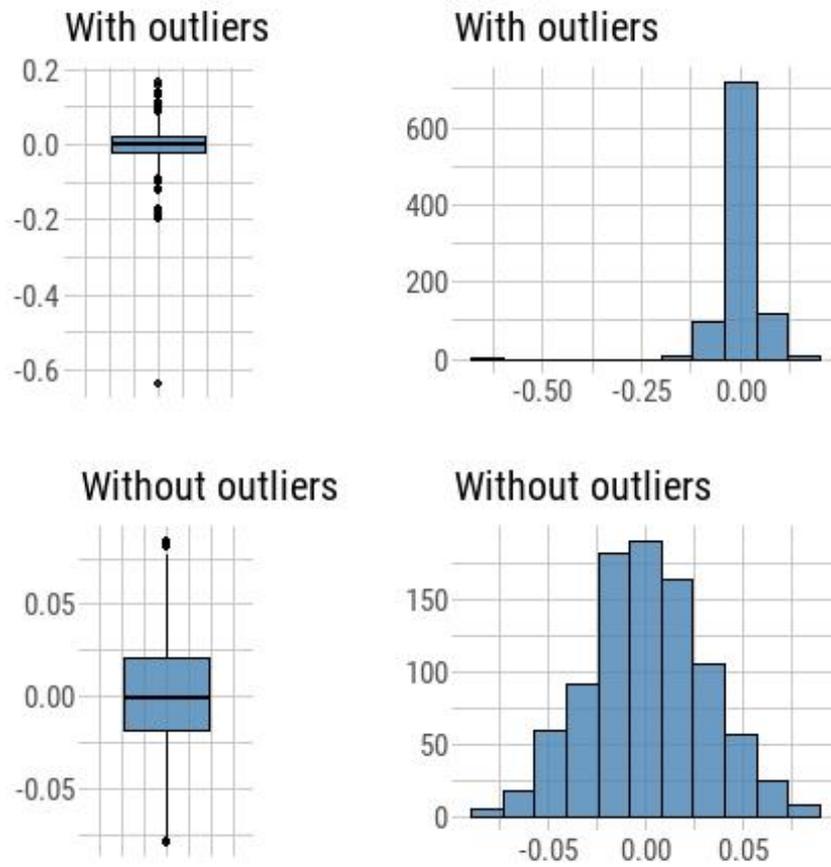


Figura 38. Outliers - Banco Galicia

Evolución de Sentimiento Diario

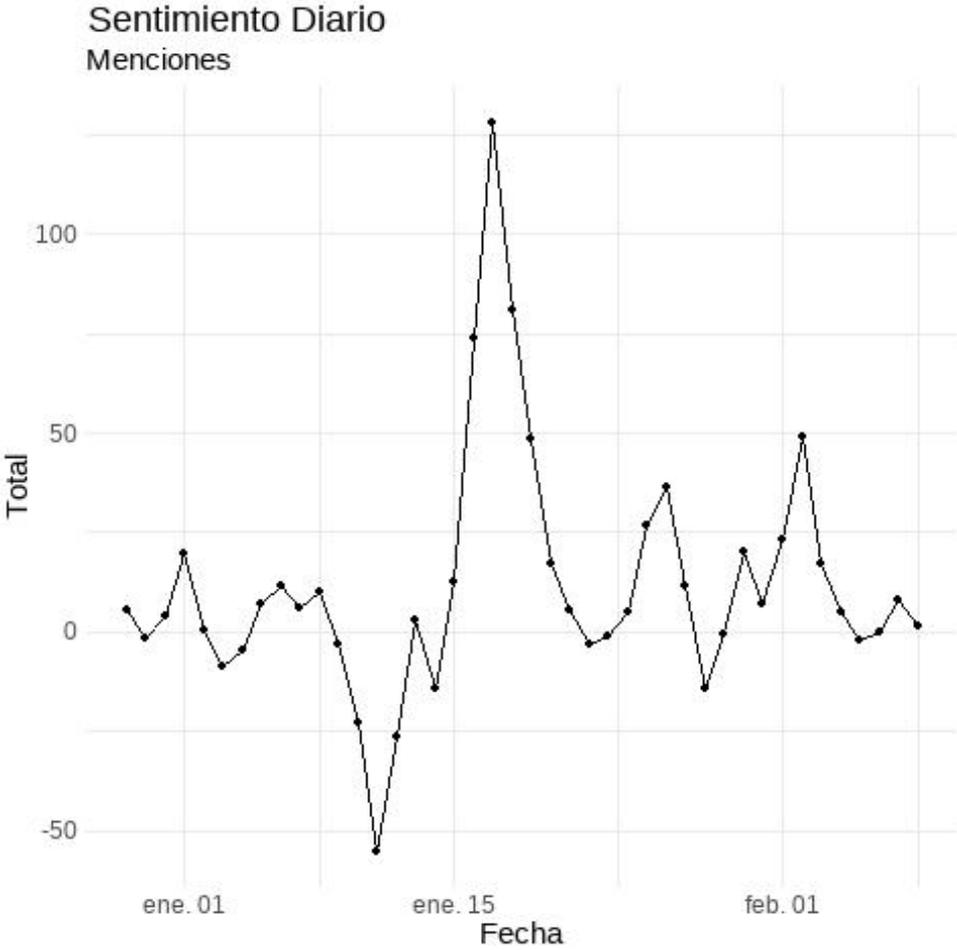


Figura 39. Sentimiento Diario de Menciones Negativas

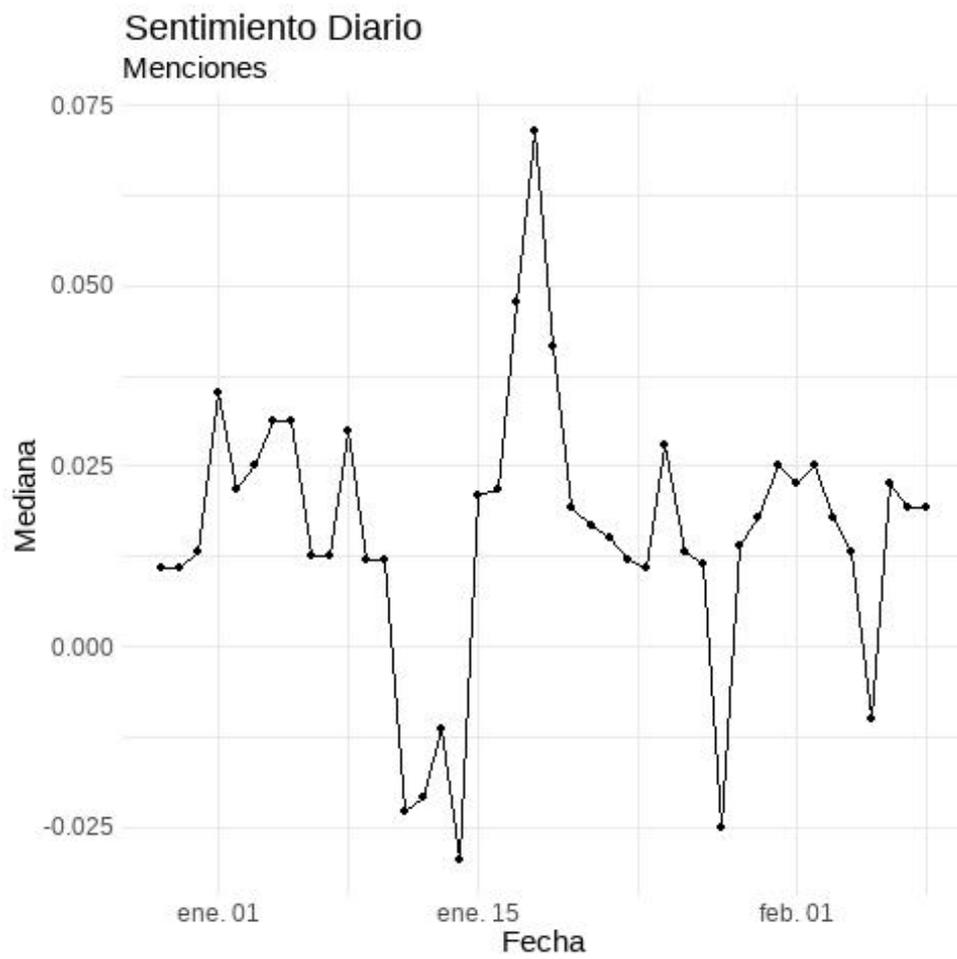


Figura 40. Sentimiento Diario de Menciones Negativas

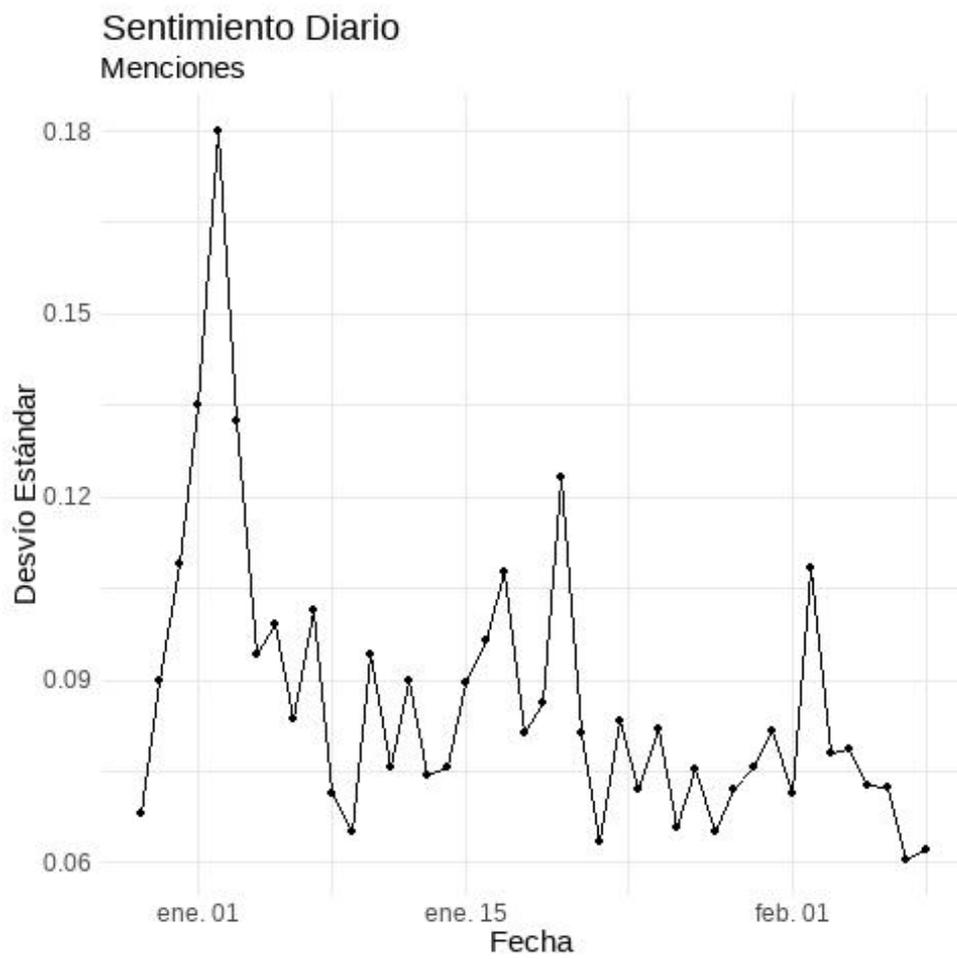


Figura 41. Sentimiento Diario de Menciones Negativas